Vol. 69, No. 4 July 1962

Psychological Review

RICHARD L. SOLOMON, Editor
University of Pennsylvania

CONTENTS

ARTICLES

Changes in the Appearance of Stimuli of Very High Luminance
TOM N. CORNSWEET 257

Practice as a Process of Simplification.....MARSHALL B. JONES 274

Normative and Ipsative Measurement in Psychology
DONALD M. BROVERMAN 295

Frustrative Nonreward in Partial Reinforcement and Discrimination Learning: Some Recent History and a Theoretical Extension
ABRAM AMSEL 306

The Selection of Strategies in Cue Learning.....FRANK RESTLE 329

Subjective Organization in Free Recall of "Unrelated" Words
ENDEL TULVING 344

The Acquisition of Knowledge.....ROBERT M. GAGNÉ 355

On the Relation between Logic and Thinking.....MARY HENLE 366

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Editorial Consultants

JOE K. ADAMS
SOLOMON E. ASCH
JOSEPH V. BRADY
JACK W. BREHM
ROGER W. BROWN
ROBERT R. BUSH
RAYMOND B. CATTELL
JAMES DEESE
EDWIN A. FLEISHMAN
ELEANOR J. GIBSON
ALBERT E. GOSS
NORMAN GUTTMAN
CALVIN S. HALL
HERBERT HOCHBERG
LLOYD G. HUMPHREYS
LEO M. HURVICH
LYLE V. JONES
HOWARD H. KENDLER
WILLIAM KESSEN
WOLFGANG KOHLER
DOUGLAS HOWARD LAWRENCE
GARDNER LINDZEY
R. DUNCAN LUCE
DAVID T. LYKKEN
IRVING MALTZMAN
GEORGE MANDLER
WALTER MISCHEL
O. HOBART MOWRER
CONRAD G. MUELLER
BENNET B. MURDOCK
ALBERT PEPITONE
LEO J. POSTMAN
MALCOLM G. PRESTON
FRANK RESTLE
SEYMOUR B. SARASON
LARRY STEIN
PHILIP TEITELBAUM
WARREN S. TORGERSON
LEONA E. TYLER
DELOS D. WICKENS
JULIUS WISHNER

(A large group of psychologists has kindly agreed to assist in the editing of manuscripts submitted to the *Psychological Review*. In each issue the Editor will list in the *Review* those editorial consultants who have helped him process manuscripts for that issue, so that their efforts may be gratefully acknowledged.)

The *Psychological Review* is devoted to articles of theoretical significance to any area of scientific endeavor in psychology. Except for occasional articles solicited by the Editor, manuscripts exceeding 7,500 words will not be accepted. Ordinarily manuscripts which consist primarily of original reports of research should be submitted to other journals. However, reports of original research may be included in a major theoretical article which attempts to integrate several related original studies.

Because of the large number of manuscripts submitted, there is an inevitable publication lag of several months. Authors may avoid this delay if they are prepared to pay the costs of publishing their own articles; the appearance of articles by other contributors is not thereby delayed.

Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstract may be obtained from the Editor or from the APA Central Office.

All manuscripts must be submitted in duplicate.

Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors should retain a copy of the manuscript, in case there is loss in the mail. Manuscripts should be addressed to the Editor, Richard L. Solomon, Psychology Department, College Hall, University of Pennsylvania, Philadelphia 4, Pennsylvania.

HELEN ORR
Managing Editor

SHARON STONE
Editorial Assistant

PUBLISHED BIMONTHLY BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Prince and Lemon Streets, Lancaster, Pa.
1333 Sixteenth St. N.W., Washington 6, D. C.

$10.00 volume $2.00 issue

Second-class postage paid at Lancaster, Pa.

Send all communications, including address changes, to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must arrive by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

VOL. 69, NO. 4 JULY 1962

PSYCHOLOGICAL REVIEW

CHANGES IN THE APPEARANCE OF STIMULI OF VERY HIGH LUMINANCE [1]

TOM N. CORNSWEET

University of California, Berkeley

When quanta of light strike the retina, some of them are absorbed by the pigment molecules contained in the visual receptor cells, and this absorption ultimately results in the activation of the visual nervous system. Recent developments in the biochemistry of the visual pigments has given us a fairly clear picture of the microstructure of quantal absorption. A theoretical analysis of this first stage in the visual process will be presented and this analysis will be shown to furnish a possible explanation of certain striking changes in hue and brightness that occur when a stimulus of very high luminance is steadily viewed.

THEORY OF VISUAL PHOTOCHEMISTRY

There is fairly good agreement that light triggers neural events by processes of the following general type. Quanta of light, after passing through the other media of the eye, arrive at the retina where they may or may not hit receptor cells. Those quanta that finally arrive in the region of the visual pigment of the receptor cells may or may not be captured by the molecules of pigment. If a quantum is in the close vicinity of a particular pigment molecule, the probability that it will actually be captured by the molecule depends upon the relationship between the wave length, or energy, of the quantum and the physical structure of the molecule and its surroundings. If all these probabilities are satisfied, the quantum is absorbed by a visual pigment molecule. When, before absorption of the quantum, the pigment molecule was in a state that will hereafter be called the "regenerated" state, the absorption will cause the molecule to change its state (to what will be called an "unregenerated" state). For example, Kropf and Hubbard (1958) suggest that the change in state is a cis-trans isomerization. The changing of state of a regenerated pigment molecule sets up a process in the receptor which eventually results in the discharge of a packet of some chemical mediator at a synapse farther along in the retina. When the concentration of chemical mediator at the synapse is great enough, a nerve impulse is triggered, and either that impulse or another

[1] This research was supported by United States Public Health Service Grant Number B2169. The writer also wishes to thank Robert Abelson and Norman Anderson for their helpful suggestions, and James Thomas for his work on the brightness reversal experiment.

resulting from it, travels up the optic nerve. At the synapse, there is also a substance that destroys the chemical mediator and this, combined with diffusion effects, give the mediator a short "half life." Therefore the probability of triggering a nerve impulse depends upon the *rate* of discharge of packets. Above threshold, the frequency of nerve impulses will be greater, the greater is the rate at which packets of mediator are being discharged. Since each discharged packet is the ultimate result of the absorption of a quantum by a regenerated molécule, the frequency of nerve impulses will depend upon the rate at which quanta are absorbed by regenerated molecules.

The changes that occur in retinal photopigment when light is absorbed are reversible, and, if the eye remains in darkness long enough, virtually all of the pigment will have regenerated to its initial, ready-to-trigger, state. Some of the more recent evidence on cone pigments *in vivo* indicates that at least the main part of this regeneration process goes on at the same rate whether the eye is in darkness or in light, the rate depending only upon the concentration of unregenerated pigment present (Rushton, 1958).

If the above general conceptions are correct, then, when a subject is steadily fixating a field of fixed luminance, retinal pigment is both being broken down and regenerated, and the frequency at which nerve impulses are being fired in any one optic nerve fiber depends upon the rate at which regenerated molecules are being broken down. Hereafter, the term "output" will be used to denote the instantaneous concentration of chemical mediator at the receptor terminus. This receptor output is thus directly related to the rate at which quanta are absorbed by regenerated molecules,

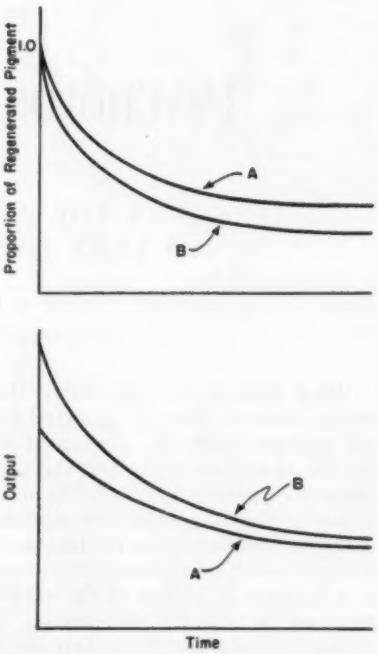

FIG. 1. Proportion of regenerated pigment as a function of the time after onset of the stimulus, for two different illuminances (B greater than A); and output as a function of the time after onset of the stimulus for the same two illuminances.

and it is similarly related to the frequency of triggering of nerve impulses somewhere farther along the retinal activity chain.

The upper curve, labeled A, in the upper part of Figure 1 is a generalized representation of the changes in pigment concentration that would occur during stimulation of a single receptor if the preceding considerations are correct. The receptor is completely dark adapted before the onset of the stimulus, so that the proportion of regenerated to total pigment is 1.0. At the time represented at the origin, a steady light is turned on. The pigment concentration will drop rap-

idly at first and then level off as it approaches some asymptotic level, the level at which the rates of breakdown and regeneration are equal. The lower curve, labeled B, represents the same events for an identical receptor under a more intense stimulus.

The curves in the lower part of Figure 1 are a generalized representation of the outputs (rates of pigment breakdown) of the same two receptors under the conditions of the upper part of Figure 1. Since the receptors are taken to be identical, and since at time 0 both concentrations are 1.0, the initial output of B will be greater than the initial output of A. Further, since the time constants of regeneration for A and B are taken to be equal, the output of B must approach a final level that is higher than the asymptote of A. This is true even though the equilibrium concentration of regenerated pigment is lower for B than for A.

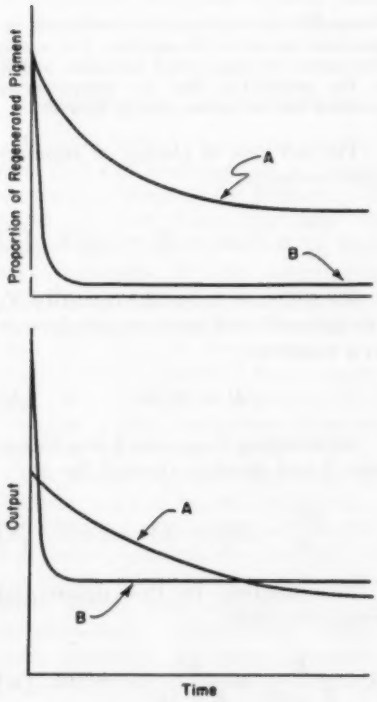

FIG. 2. Proportion of regenerated pigment as a function of time after onset of the stimulus, for two different illuminances (A at moderate illuminance and B approaching an infinite illuminance); and output as a function of time after onset of the stimulus for the same two illuminances.

The upper part of Figure 2 represents the regenerated pigment concentration curves for the same two receptors when the illuminance on B has been raised almost to infinity, but the illuminance on A is the same as in the preceding figure. In the first short interval after the light is turned on, virtually all of the pigment molecules in B will be broken down. From that time on, each molecule that flips back to the regenerated state will immediately be broken down again. In other words, when the illuminance approaches infinity, the concentration approaches equilibrium extremely rapidly, and the equilibrium level depends only upon the time constant of regeneration of the pigment. The lower part of Figure 2 shows the output curves for these receptors, again when the illuminance on B approaches infinity. It will again be true that, when the stimuli are first turned on, B will have the higher output. Further, at equilibrium, B must still have the higher output, even though its regenerated pigment concentration is very much lower. In general, the equilibrium rate of breakdown of regenerated pigment (output) is greatest when the illuminance is infinity, since that rate equals the rate of regeneration and since the rate of regeneration will be greatest when the concentration of *unregenerated* pigment is greatest. But, since

Receptor B will reach its equilibrium level almost instantaneously, its output curve must cross the output of A twice, as in the lower part of Figure 2. (This is necessarily true as long as the illuminance on A is great enough that its initial output level is greater than the equilibrium level for B.) Obviously the illuminance on one of a pair of receptors need not be nearly infinite in order that the pair exhibit a reversal of output relations. There will be a large set of pairs of illuminances for which reversal should occur.

A More Specific Set of Assumptions and Their Consequences

Now let us substitute, for each of the general conditions described above, a corresponding specific assumption about the operation of the visual system. From these assumptions, certain quantitative aspects of the visual phenomena to be discussed may be derived, and the derivations examined in comparison with the data themselves.

If the absorption of a quantum by a particular molecule is independent of the state of the other molecules in the receptor, then it follows from statistical considerations that the rate of absorption by regenerated molecules is:

$$\left(\frac{dN}{dt}\right)_- = AQN \qquad [1]$$

where N is the number of regenerated molecules, A is the probability that a quantum incident on a region containing a regenerated molecule will bleach that molecule (the photosensitivity of each regenerated molecule at the wavelength of the incident quantum), and Q is the number of quanta incident during the Time dt.

Recent studies of the bleaching of cone pigments in the living human eye support the notion that, at least to a good approximation, the regeneration of cone pigments is a first order chemical reaction (Rushton, 1958). We will assume, accordingly, that the probability that any particular unregenerated molecule will be regenerated in a unit of time will be independent of the states of the surrounding pigment molecules. It follows then that the rate of regeneration of unregenerated molecules is:

$$\left(\frac{dN}{dt}\right)_+ = K(M - N) \qquad [2]$$

where M is the total number of molecules, regenerated or not, in the receptor, N is, again, the number of regenerated molecules, and K is the probability that an unregenerated molecule will be regenerated in Time dt.

The net rate of change of regenerated molecules is:

$$\frac{dN}{dt} = K(M - N) - AQN \qquad [3]$$

We will now define the quantity X, the proportion of regenerated pigment in a receptor:

$$X = N/M \qquad [4]$$

Substituting Equation 4 into Equation 3 and dividing through by M:

$$\frac{dX}{dt} = K(1 - X) - AQX \qquad [5]$$

The solution to this differential equation yields:

$$X = \frac{K}{K+AQ} + \frac{AQ}{K+AQ} \cdot e^{-t(K+AQ)} \qquad [6]$$

when the initial condition is $X = 1$, that is, complete dark adaptation.

In the preceding section, the output of a given receptor was defined as the rate at which quanta are absorbed by

regenerated molecules. Therefore:

$$\Omega \equiv \left(\frac{dN}{dt}\right)_{-} = AQN = AQMX$$

$$\Omega = AQM\left[\frac{K}{K+AQ} + \frac{AQ}{K+AQ}\cdot e^{-t(K+AQ)}\right] \quad [7]$$

This is an equation that relates the output of any given receptor to the irradiance and duration of a steady light falling upon it.[2] The lower parts of Figures 1 and 2 are actually plots of Equation 7 for various values of Q.

CONSEQUENCES FOR BRIGHTNESS

It seems to be a reasonable assumption that the brightness of a lighted area will increase as the frequency of nerve impulses originating from the retinal region corresponding to that area increases. If this assumption and those concerning the conversion of quanta to impulses are correct, then the falling of the curves in the lower parts of Figures 1 and 2 represent the fact that the brightness of a stimulus of fixed luminance decreases as the eye light adapts to it.

The considerations illustrated in the lower part of Figure 2 also lead to the following prediction. If the eye is first dark adapted, then one area of the retina is illuminated (say with white light), and another area is simultaneously illuminated much more intensely, the more intensely illuminated area will look first brighter than the other, then darker, and finally somewhat brighter again. That is, there should be a period of time during which the less intensely illuminated region should look brighter than the more intensely illuminated one. This paradoxical prediction has been tested and it is correct. The predicted brighter reversal does indeed occur, and is very easy to demonstrate.

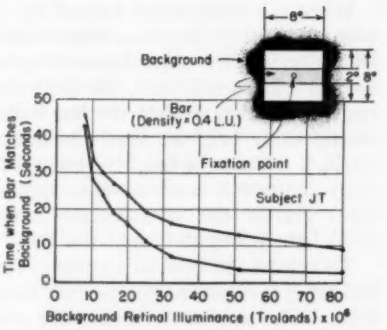

FIG. 3. Times of brightness reversals as a function of retinal illuminance for the stimulus configuration shown in the upper right corner. (The surround was dark, the background illuminated, and the horizontal bar consisted of a strip of 0.4 log unit neutral density filter. Each point is a mean of four judgments.)

The data in Figure 3 were taken under the following conditions. The subject fixated a small bright point in the center of the field shown in the upper right corner of the figure. The field was shown in Maxwellian view, through a simple optical system whose effective aperture was 3.5. The source was a 32-candle power automobile taillight bulb, run from a regulated DC supply. The filament image at the pupil was 1 millimeter in diameter, centered on the natural pupil. The luminance of the field was varied with neutral density filters. Heat glass was used to filter out the infrared.

[2] None of the steps in this derivation is original. Pieces of it can be found in the writings of Troland (1930), Hecht (1935), Rushton (1958), for example. And, although the writer has not found the complete derivation all in one place in the literature, it may well be there somewhere.

When the field is first turned on, it looks like a dark bar on a bright field. But after a time that depends upon the luminance relations, the field abruptly reverses, so that the bar looks considerably brighter than the background. Then the brightnesses gradually converge and reverse again, until finally the background looks just slightly brighter than the bar.

To collect the data in Figure 3, the following procedure was used on each trial. The subject's viewing eye was dark adapted for 5 minutes. Then the stimulus figure was turned on and two clocks started. The subject looked steadily at the fixation point, and at the first instant when the bar and its background were equally bright (changing from normal to reversed brightnesses) he pressed a key, stopping one of the clocks. When the brightnesses were again equal (changing from reversed to normal) he pressed a second key, stopping the second clock. The two times were recorded, and each point on the curves in Figure 3 is a mean of four such times. The area between the two curves in Figure 3 represents the times when the brightnesses are reversed. At least 45 minutes elapsed between successive trials, and the order of illuminances was balanced.

The second equality point is much harder to judge than the first one, and the judgments are less reliable. Part of the reduction in reliability is due to the fact that the changes are occurring very gradually, but, in addition, strong hue differences between the bar and ground appear after the first 20 seconds or so; the background turns reddish and the bar less red. (There is not much point in trying to interpret these color differences physiologically, because the wave length composition of the stimuli in this set of experiments

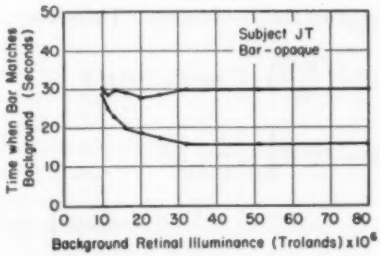

FIG. 4. Times of brightness reversals as a function of retinal illuminance. (For these data, the bar, as shown in Figure 3, was replaced by an opaque strip. Each point is the mean of four judgments.)

was very complex, due to absorption by the heat glass, etc.)

The brightness reversal occurs equally strongly whether the fixation point is in the bar or the background. It cannot be explained by inhomogeneities in the retinal structure.

The data in Figure 3 were taken when the bar was .4 log units less intense than the background. (The bar was simply a strip of .4 density neutral gelatin filter pasted across the larger square aperture that formed the background.) But brightness reversal also occurs when the bar transmits no light to the retina. The data for Figure 4 were taken after the strip of .4 density filter was replaced by an opaque strip. In this case, the time relations are less strongly dependent upon the illuminances (and the hue differences more extreme), but the effect is still very strong. The fact that the strip is opaque does not mean that the actual retinal illuminance in the region of the image of the bar is zero. There is extensive evidence in the literature to indicate that the stray light in this sort of stimulus situation is appreciable (see DeMott & Boynton, 1958, for an extreme example). Furthermore, there is a good possibility that the illuminances at

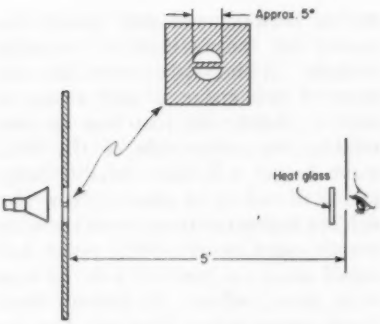

FIG. 5. Apparatus for a demonstration of brightness reversal. (The inset figure shows the subject's view.)

the edges of the figure (where stray light is greatest) are the critical ones. (This possibility is currently being tested.) Thus, the retinal condition in which an opaque bar is used in the apparatus is on a continuum with any in which a filter of transmittance greater than zero is used.

It is easy to set up a demonstration of this brightness reversal, as illustrated in Figure 5. A strip of cardboard is placed across a round hole in a larger board. This hole is just a little smaller than the face of the sort of photoflood lamp that has a frosted face and a built-in reflector (e.g., GE #PH/RFL2). A small hole in the middle of the strip will serve as a fixation point (it is essential to fixate very steadily for this demonstration). The face of the photoflood lamp should be viewed through the hole and strip, and through heat absorbing glass, from about 4 or 5 feet away.

AN EXTENSION OF THE THEORY TO CHANGES IN HUE

When a subject steadily fixates a field of extremely high luminance in the long wave length region of the spectrum, the hue of the field undergoes very striking changes. Some of these changes have been mentioned independently by Auerbach and Wald (1955) and by Noddack and Jarzyk (1955), in the course of studies of other color phenomena. Cornsweet, Fowler, Rabedeau, Whalen, and Williams (1958) briefly reported a study of the color changes themselves. According to all of those articles, the nature of the color changes is as follows. The subject looks steadily at a field that is illuminated with light from the long wave length end of the spectrum. When the luminance is low or moderate, the field will look red or orange. However, if the luminance is extremely high, it will look orange at first, but then rapidly shift in hue through yellow to a vivid green. This pattern of changes is reported to occur for stimuli of several different wave lengths in the "red" end of the spectrum, and for monochromatic as well as broad band stimuli. Stimuli with wave lengths in the green region of the spectrum are reported to produce extreme desaturation, but no clear hue changes are reported. Auerbach and Wald (1955) also report that bright blues will change to red, but such a change was not observed in either of the other two studies cited. A review article by Cohen (1946) provides a number of references to other studies of the effects of high luminance stimulation.

The theoretical considerations discussed in the preceding sections suggest an explanation for some of these changes in hue. The explanation is based upon the retinal mechanisms already mentioned, along with some simple and widely held assumptions about the physiological bases for color vision.

First, we will assume that there are at least two sorts of cones that differ in the pigments they contain, the two pigments having different but overlapping absorption spectra. The absorption spectra for such a pair of

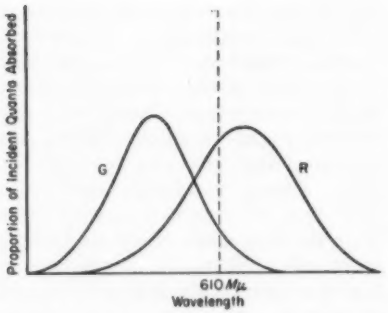

FIG. 6. Generalized spectral absorption curves for two hypothetical retinal pigments. (The wave length 610 mu, indicated on the abcissa, will be specifically referred to later in the text.)

cone types are shown in Figure 6.[3] The two cone systems will be labeled R and G, as in Figure 6. Now consider the situation in which a subject is shown a stimulus and is forced to call it either red or green, no other alternative being permitted. Let us assume that the probability of his giving a "red" judgment depends upon the ratio of the outputs of the R and G systems. In the simplest case, when the R output is greater, he will be more likely to call the stimulus red, when the G output is greater, he will call it green, and when they are equal, he will judge it green half the time and red the other half. There is no theoretical necessity that the output ratio be 1.0 for 50% "red" judgments, but that particular condition will be assumed for much of the remainder of this paper, merely to make the discussion easier to phrase.

Figure 7 shows the results obtained when a subject is forced to judge a flash of monochromatic light as either red or green, as the wave length in the region between red and green was varied by the method of constant stimuli. It should be noted that the slope of this curve is very steep, so that a change of just one or two millimu on either side of the 50% wave length will elicit reliable judgments of red or of green. Now if a subject is given a stimulus with a wave length right at the 50% point and asked what its hue *really* is, he says it is pure yellow. It follows from these observations that the last assumption discussed above is equivalent to assuming the following: when the R output exceeds the G output, the subject will see red, when the G output is greater, he will see green, and when the two are equal, he will see yellow. (This is of course true only for the spectral region between red and green. There may be other spectral regions in which the R and G outputs are also equal, but where the hue is given some other name, e.g., blue. The question of whether

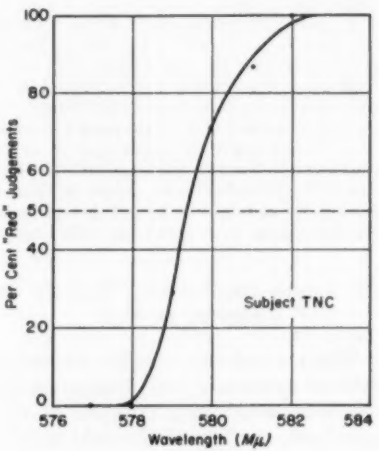

FIG. 7. Distribution of judgments for constant stimulus method. (On each 1/4-second flash, the subject was given a forced choice between responses "red" and "green.")

[3] The theory discussed in this paper requires that there be *at least* two cone systems, but does not approach the question of how many there may be altogether.

the yellow sensation is produced by a mixture of the R and G outputs or represents an excess of yellow over blue output with the red and green outputs cancelling each other, is not relevant to any of the present discussion.)

The explanation of some of the hue changes that occur to stimuli of high luminance follows directly from these assumptions. Monochromatic light in the long wave region of the spectrum, for example at 610 millimu, will be absorbed by both the R and the G cone pigments, but since the absorption coefficient for the R pigment is higher than the G at 610 millimu (see Figure 6), the R cones will absorb a greater percentage of the incident quanta. This is exactly equivalent to putting a greater illuminance on the R than the G cones. Therefore, if the R and G cones differ only in their absorption spectra, and a stimulus of very high luminance at 610 millimu is given, their output versus time curves will look like those in the lower part of Figure 2, Curve A representing the G and B the R system. The stimulus should look red at first, then change through yellow (when the curves cross) to green, and finally slowly change back through yellow to orange.

It has been reported in previous papers (Cornsweet et al., 1958) that a high-luminance red stimulus will appear to change from red through yellow to green. "No further color changes were observed, even though fixation was maintained for as long as three minutes" (p. 898). The present more extensive data at a variety of wave lengths, and at luminances greater than those used in the earlier study, clearly show that further changes do occur. In fact, the hue of the stimulus eventually changes back from green through yellow to orange as is predicted by the theoretical argument above. This later set of changes happens to take appreciably longer than 3 minutes at the relatively low luminances and at the particular wave lengths used in the earlier study (Cornsweet et al., 1958). But at other wave lengths and luminances, all of the changes appear to be over within a minute or two (see the data in succeeding sections).

General Aspects of Hue Changes

The general course of the hue changes that occur when viewing a high-luminance stimulus at 620 millimicrons is illustrated in Figure 8. This figure is drawn from a set of crude binocular matchings, in which the high-luminance stimulus was presented to one eye and a matching stimulus of much lower luminance, whose wave length could be varied, was presented to the other. The curve in this figure should not be taken as anything more than a qualitative indication of the phenomenon under discussion. Quantitative data of a more reliable and useful sort will be presented later.

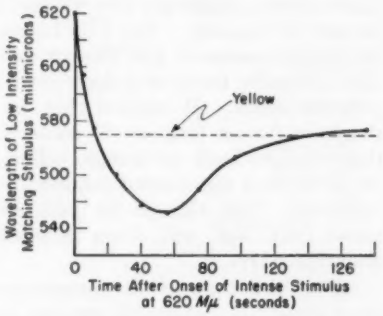

FIG. 8. Wave length of a low luminance stimulus flashed to the left eye that matches the hue of a high luminance stimulus, at 620 mu, steadily fixated by the right eye, as a function of the time of viewing the high-luminance stimulus.

At this writing, the limits of the range of wave lengths for which hue changes occur have not been well established. However, the following general statements are true. High-luminance stimuli with wave lengths in the "green-yellow" region of the spectrum (from about 560 to 580 millimu) look green at first, turn a much deeper green, then back through yellow to orange, and finally back to yellow or greenish yellow. Any wave length between 580 and 655 mu will look orange at first, change through yellow to a good, saturated green, and finally back to an orangish yellow. It is probable that light of wave lengths longer than 655 mu also appears to turn green if its luminance is high enough, but 655 mu is the limit with the present apparatus.

Stimuli below 560 mu have not been studied systematically, but in general they seem simply to desaturate to a white at the luminances available in the present apparatus.

For stimuli in the region of about 620 down to 560 mu or less, when the luminances are well above those necessary to produce the change to green, another change seems to be superimposed upon the first 15 or 20 seconds of viewing. The field begins its normal course of hue change, but then abruptly turns to a desaturated pink or violet. It retains that appearance for as long as 20 seconds, then changes back to orange, yellow, or green, and the normal sequence is resumed. This change to pink or violet has not yet been studied quantitatively.

All of the changes in appearance that have been observed depend in character and timing upon the luminance, and, of course, the wave length of the stimulus. These dependencies will be analyzed below.

QUANTITATIVE ANALYSIS OF HUE CHANGES

The photochemical theory outlined above requires that the time course of changes in hue be systematically related to the luminances and wave lengths of the stimuli. In order to evaluate the theory quantatively, data have been collected on the time course of hue changes as a function of retinal illuminance for a number of wave lengths. Those data and their relationship with the theory will be presented below.

PROCEDURE

Judgments Used for the Collection of Quantitative Data

The data presented here were collected by using two different sorts of procedures. The data for Figure 8 (and the data for the brightness reversals discussed earlier) are based upon a matching technique. Pairs of stimuli are found which differ physically but are judged the same along some perceptual dimension (hue in Figure 8, hue and brightness for the brightness reversal data). The other procedure used here is a sort of hue naming or identification. For example, most of the data in this report result from recording the times at which a subject judges the hue of the stimulus to be pure yellow, that is, neither reddish yellow nor greenish yellow.

In situations where a choice may be made between matching and hue naming techniques, it is usually far preferable to employ matching (Brindley, 1960). Hue naming was used principally here for two reasons. First, for the particular phenomenon studied, the problems introduced by matching are considerably worse than those of hue naming. For example, to determine the time at which a stimulus of high luminance produces a hue match with a standard hue (e.g., yellow) the matching field must be at a low enough luminance that its hue is not changing very much. Therefore, the luminance of the matching field should be at least 2 log units lower than of the test field. However, the stray light from the test field then approaches and perhaps even exceeds the direct illuminance of the matching field. If the matching field is presented to one eye and the test field to the other, the

effects of the stray light are not eliminated, but just rendered less estimable. The collection of the data represented in Figure 8 took advantage of the fact that, even when the two fields are projected on to noncorresponding parts of the two retinas, rivalry occurs. That is, if the stimulus of high luminance is presented continuously and the matching field is occasionally flashed to the other eye, the only thing visible during the flash seems to be the matching field. It is then possible to find a test flash wave length that is the same hue as the high-luminance field at a given time after the onset of the high-luminance field. However, this procedure clearly involves so many factors of unknown effect (stray light, rivalry, transients from the flash, etc.) that it is at best useful only to give a rough idea of the sorts of hue changes that do occur.

A different sort of matching technique might have been used in this study. Instead of determining the time at which the intense field matches one of a fixed hue, the subject might have been presented with two patches having different wave lengths and high but different luminances. Then as the patches both changed in hue (and brightness), the time when they match could have been determined. While not all pairs of stimuli would reach a match, there must be sets of them that would do so. This kind of matching procedure would not be so severely affected by stray light, and in many ways is preferable to the color naming procedure. Brindley (1960) has clearly pointed out that this type of matching procedure (or at least one almost identical with it) enables one to draw physiological conclusions from psychophysical data, without requiring what he calls "psychophysical linking hypotheses." However, it is also true that, when such a procedure is used, psychophysical linking hypotheses *cannot* be tested, and one of the aims of this paper is to examine just such a question, namely, what are the physiological correlates of the condition under which a subject reports the hue as red half the time and green the other half. The matching technique however does have many advantages over color naming as long as psychophysical linking hypotheses are not in question, and studies using strict matching procedures are now in progress.

On the other hand, the procedure that consists of recording the instants in time when the subject calls the stimulus neither reddish nor greenish yields very reproducible results, and results which may be interpreted in a relatively straightforward way.

Apparatus

The apparatus used to collect all of the color data in the present study is shown schematically in Figure 9. The source, S, was a tungsten-coil-filament lamp (GE PH/18A/T10P), run from a constant voltage transformer at 6 volts, 18 amperes. Lens L_1 formed an image of the filament on the entrance slit, En, of a Bausch and Lomb grating monochromator of 500 millimeter focal length, with a 600 lines per millimeter grating, blazed in the visible. The optics of the monochromator imaged the entrance slit (and the filament) on the exit slit, Ex, with an equivalent aperture ratio of 4.4. An airspaced, coated, achromatic lens, L_2, focal length 150 millimeters, f = 1.9, imaged the exit slit in the plane of the pupil of the subject's eye so that he saw a Maxwellian view of the back surface of the lens L_2. The entrance and exit slits of the monochromator were set to widths of 1.5 millimeters, resulting in stimuli with spectral band widths of 5 millimu. The

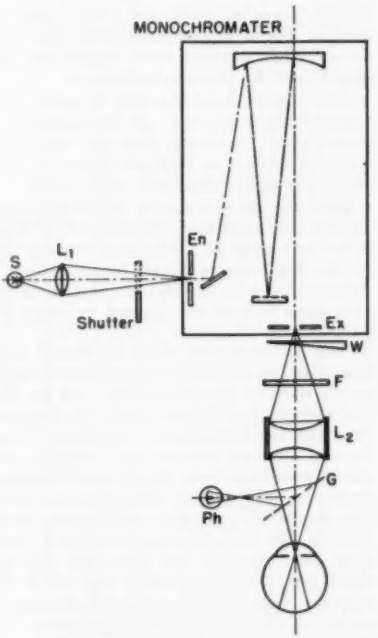

FIG. 9. Apparatus for delivering very intense monochromatic stimuli. (See text for description of specific elements.)

length of the exit slit was also set to 1.5 millimeters. Therefore, since Lens L_2 gave unit magnification, the image at the pupil was a square 1.5 millimeters on a side. This is small enough that as long as the exit slit is imaged at the center of the natural pupil, changes in the size of the natural pupil will not produce appreciable changes in retinal illuminance. A neutral-density wedge, W, and filters, F, were used to adjust the luminance of the stimuli. The wedge was close enough to the exit slit that no balancing wedge was needed. A clear piece of glass, G, reflected a small fraction of the stimulus light to a photomultiplier tube, PH, and the output of the phototube was used to monitor the luminance of the stimulus.[4]

A Hilger thermopile and mirror galvanometer were used to make the energy calibrations. The thermopile was placed at the pupil position during these measurements. The spectral purity of the stimulus lights was measured with a second monochromator set at a 5 millimu band width. When the stimulus was at any of the wave lengths studied, no other band had a luminance as great as 1% of the stimulus luminance, and the total light at other wave lengths was less than 5% of the stimulus luminance.

The subject's head was held in position by a dental-impression bar. At his hand was a control panel containing push buttons which operated timers, so that the timing of any changes in appearance could be recorded.

When the stimulus was on, the subject saw a circular field, uniformly illuminated, as illustrated in Figure 10. The dark circular line in this field was produced by cementing a circle of fine wire to the back surface of L_2. This circle helped the subject to make his judgments, in that he judged only the hue within the circle, ignoring the occasional edge effects that small eye movements produced at the outer margins of the field.[5]

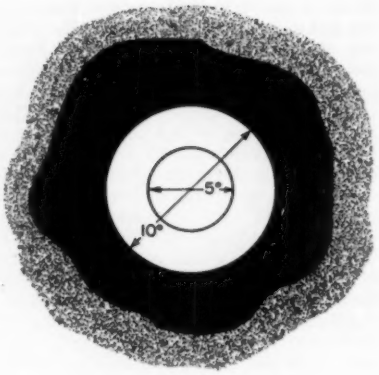

FIG. 10. Stimulus pattern as it appeared to the subject. (The entire 10-degree circular area was uniformly illuminated with monochromatic light. The subject was instructed to fixate the center of the pattern and judge the hue of the inner disk.)

RESULTS OF HUE CHANGE STUDY

The data plotted in Figure 11 show how the time taken to turn yellow varies as a function of the retinal irradiance of a monochromatic stimulus at 610 millimu. For example, at an irradiance of 3×10^{12} quanta per second per square millimeter of retina (uncorrected for ocular transmission losses), the patch looked red or orange at first. The subject pressed his key to indicate that it was yellow (neither reddish nor greenish) after 17 seconds of viewing. He pressed a second key to indicate the second occurrence of

[4] For the relatively short slit-length used here, the coil filament bulb delivered more energy through the system than did the ribbon filament bulb that is ordinarily supplied with the monochromater. However, using the coil filament, the energy at the exit slit varied somewhat from day to day in an unpredictable way. This variability was probably due to the fact that only a very small piece of the filament image was used, and, since the image is not uniformly intense, slight amounts of sagging of the coil or its supports could result in noticeable changes in output at the exit slit. For this reason, the monochromater output was continuously monitored, and the intensity settings were made from the photomultiplier output rather than from the wedge or filter calibrations.

[5] The particular choice of a 10-degree stimulus disc is not crucial. The hue changes occur with fields as small as 15 minutes of arc in diameter. (We have not tested still smaller sizes.) But the variance of the temporal aspects of the judgments is smallest when the field is large.

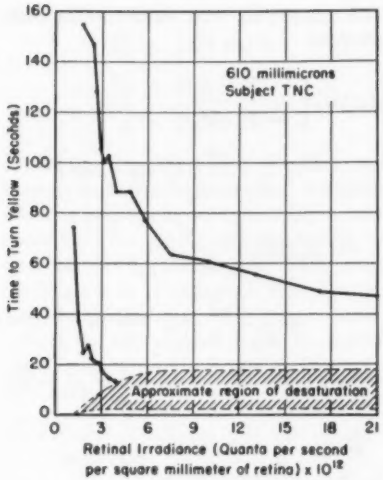
610 millimicrons
Subject TNC
Approximate region of desaturation
Time to Turn Yellow (Seconds)
Retinal Irradiance (Quanta per second per square millimeter of retina) x 10¹²

FIG. 11. Time to change hue as a function of the retinal irradiance of the stimulus, at 610 mu. (Each point is a mean of two readings.)

yellow 100 seconds after the stimulus had first been turned on. The area between the two curves represents the times during which the stimulus looked green.

The upper curve, for the second change to yellow, extends out to a retinal irradiance of 21×10^{12} quanta per second per square millimeter, the maximum irradiance available in the apparatus. The lower curve is not plotted beyond 3.9×10^{12} quanta/sec/mm^2 because the hue changes in that region were obscured by a change to desaturated pink or violet, as indicated roughly on Figure 11. For this wave length, the stimulus never appeared yellow at irradiances below 1.2×10^{12} quanta/sec/mm^2.

Figure 12 shows the same sort of data for light of three other wave lengths in the long wave end of the spectrum. The curves for 600 millimu are very similar to those for 610 millimu (Figure 11). As the wave length increases, the lower curves shift toward the higher irradiances, but retain the same general shape. The upper curve at 630 millimu shows evidence of a peak at the lower retinal irradiances. The upper curve at 640 millimu exhibits a very large amount of variability, and is not plotted beyond 9×10^{12} quanta/sec/mm^2 because the change back to yellow had not occurred during 210 seconds of continuous viewing. The variability in all cases increases very greatly as the time of viewing becomes extremely long. It will be shown that curves of this form may be expected from the theoretical considerations discussed previously.

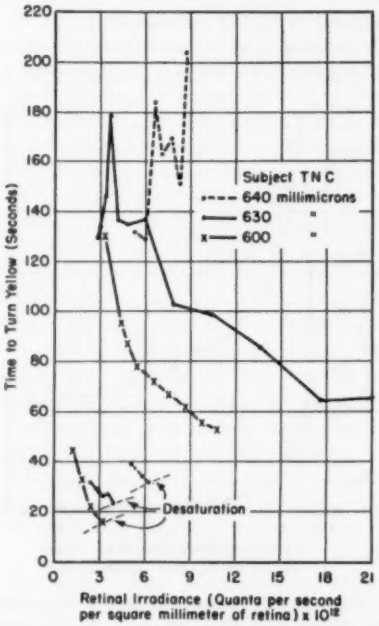

FIG. 12. Time to change hue as a function of retinal irradiance for three different wave lengths. (Each point is a single judgment.)

RELATIONSHIP BETWEEN THE DATA AND THE THEORY

The data in Figure 11 have been replotted as dots in Figure 13. They show how the time taken to turn yellow varies as a function of the irradiance of a monochromatic stimulus at 610 millimu. Every dot in this figure represents a time when the stimulus looked yellow. Therefore, according to the theoretical arguments outlined above, each of these points stands for a condition where the outputs of the R and the G systems are in some particular ratio. That is:

$$\Omega_r = C\Omega_g \qquad [8]$$

where C is the value of the ratio at which the hue is equally often called red and green in a forced choice. (In the preceding discussion C was assumed to be 1.0 for simplicity.)

Substituting the photochemical Equation 7 for output in Equation 8,

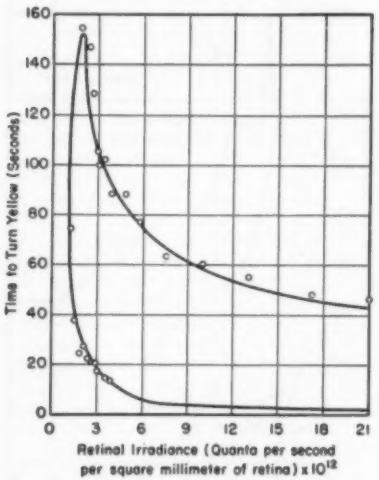

FIG. 13. Time to change hue as a function of retinal irradiance, for light at 610 mu. (Each point is a mean of two judgments. The smooth curve is a plot of the iterative solutions to the implicit Equation 10, using the parametric values listed in the text.)

and supplying the appropriate subscripts:

$$A_rQM_r\left[\frac{K_r}{K_r+A_rQ} + \frac{A_rQ}{K_r+A_rQ}\cdot e^{-t(K_r+A_rQ)}\right] = CA_gQM_g\left[\frac{K_g}{K_g+A_gQ} + \frac{A_gQ}{K_g+A_gQ}\cdot e^{-t(K_g+A_gQ)}\right] \qquad [9]$$

where t is the time of occurrence of yellow. This equation may be rewritten:

$$\frac{K_rA_r}{K_r+A_rQ} + \frac{A_r^2Q}{K_r+A_rQ}\cdot e^{-t(K_r+A_rQ)} = J\left[\frac{K_gA_g}{K_g+A_gQ} + \frac{A_g^2Q}{K_g+A_gQ}\cdot e^{-t(K_g+A_gQ)}\right] \qquad [10]$$

where $J = M_gC/M_r$.

A crude iterative procedure has been used to find the values for the five parameters of this equation that yield a good fit to the data. The two smooth curves drawn in Figure 13 are a plot of Equation 10, when the following values of the parameters are used:

time constant of regeneration of red system, $K_r = .010$ sec^{-1}

time constant of regeneration of green system, $K_g = .0083$ sec^{-1}

photosensitivity of red system at 610 mu, $A_r = 1.70 \times 10^{-15}$

photosensitivity of green system at 610 mu, $A_g = 1.32 \times 10^{-15}$

ratio of amount of green to red pigment multiplied by the output ratio for yellow judgment, $J = 1.18$.

These preliminary values may be compared with Rushton's (1958) esti-

mates:

$$K_r = .008 \text{ sec}^{-1}$$

$$K_g = .008 \text{ sec}^{-1}$$

$$A = .25 \times 10^{-15}$$

for the cone system as a whole, based upon measurements of the bleaching of cone pigments by white light, and corrected for an ocular transmission loss of 50%. Assuming that only an R and a G pigment are present, and correcting for 50% transmission losses, the estimate of A for the entire cone system on the basis of the present data is:

$$A_t = 1/.5A_{r,g} = 2[1-(1-A_r)(1-A_g)]$$

$$\approx 2(A_r+A_g)$$

$$A_t \approx 6 \times 10^{-15} \text{ cm}^2$$

The value of A_r determined from the color reversal data agrees much more closely with that quoted by Brindley (1960). From psychophysical data (quite different from those presented in the present paper), he estimates that $A_r = 1.61 \times 10^{-15}$ (p. 217). The writer has been unable to find estimates of the constant "J" in the literature. (This constant represents a psychophysical linking hypothesis, in Brindley's terminology.)

The agreement between these estimates is not perfect, nor is the fit in Figure 13. Its presentation is not intended to imply that the present formulation and these particular values are the "true" ones. The plot is meant to serve two purposes. First, it makes clear that the general formulation is consistent with the reported hue changes when put into at least this specific form. Second, it serves as a preliminary report on the extensive work now in progress attempting to evaluate, quantitatively, the specific theoretical arguments presented here. This work consists in making determinations of the best fit values of the parameters of Equation 10 under a number of independent conditions. For example, if this procedure is repeated at a number of different wave lengths in the red region of the spectrum, the best fit values of K_r, K_g, and J should be identical at each wave length, but the values of A_g and A_r should trace out the spectral sensitivity curves of the R and G systems.

The quantitative procedure used for finding parameters of the color phenomena may be applied to the achromatic brightness reversal phenomenon as well. However, the particular stimulus configuration used to gather the brightness reversal data presented in this paper do not lend themselves to straightforward analysis. For example, it is almost certain that the illuminances at the borders of a retinal region are more important in determining the apparent brightness of the region than are the illuminances in other regions; and, for a bar against a background, stray light and small eye movements have large and indeterminate effects on the effective illuminances at the borders. An experiment is currently in progress in which these factors are minimized, while the time course of brightness reversal is determined in much the same way as before. The results of this experiment should be directly analyzable in terms of the photochemical hypothesis presented above.

Observations on the rate at which the hue of the stimulus changes are qualitatively consistent with the theory, as plotted in the lower part of Figure 2. That is, the first change through yellow passes quickly, the stimulus actually appearing to be yellow for a very short time. The second change, where the theoretically determined output curves are running

more nearly parallel, occurs very slowly and gradually. This is borne out quantitatively by the fact that the variance of the judgments of the time of second yellow point is appreciably greater than the variance of the first yellow point. The brightnesses of the stimuli also agree qualitatively with the theory. Although brightness has not been measured for the monochromatic stimuli, it is always high at first, gradually dropping to a much lower equilibrium level.

From the theory as it has so far been stated, it might be expected that, just as a red light turns green, a green light should turn red, and a yellow one should not change in hue. This would indeed be the prediction if it happened to be true that (*a*) the rates of regeneration of the red and green systems were identical, and (*b*) the density of red and green pigment molecules were also identical. But if one regeneration rate were greater than the other, high-luminance stimuli near the yellow region would, at *equilibrium*, take on the hue of the system that regenerates faster. From the present findings, this hue seems to be red, but the data are not yet complete enough to draw firm conclusions about the relative rates of regeneration.

The fact that yellow and greenish yellow stimuli change toward purer greens before changing back through yellow would follow directly from the theory if there is more green than red pigment present in the stimulated region of the retina. Consider, for example, light of a wave length that looks yellow when it is first turned on. This means that it is initially isomerizing red and green pigment molecules at the same rate (again assuming that $C = 1$ for convenience). However, if there are actually fewer red molecules present per unit of area, the rate of isomerization of red molecules will immediately become smaller than the corresponding rate for the green system, and the stimulus should turn green. This sort of effect is superimposed upon the processes already discussed so that, at the proper irradiances, an initially green stimulus should first get greener, then change back through yellow to red, and finally again through yellow to green, as it does.

It was mentioned above that, when the wave length is not too long (in the "orange") and the irradiance is extremely high, the early stages of hue change seem to be obscured by a wave of desaturated pink or violet. This effect would be expected if short wave sensitive systems were present and did produce desaturation. In this case, during the early stages of adaptation, both the R and the G systems are so intensely bombarded, relative to the short wave sensitive systems, that the R and G outputs drop to values comparable with the short wave systems, and desaturation occurs.

Note: One should exercise caution when viewing stimuli of very high luminance. Fechner permanently damaged his retinae by staring at the sun, and ophthalmologists can count on an increased patient load after solar eclipses. Since there have been very few systematic studies of the pathological effects of intense lights (see Whiteside, 1960), there are no clear rules on what should be avoided. In general, the sun should not be used as a source unless it is very strongly filtered, and any source that contains a large amount of infrared radiation should be filtered by heat-absorbing glass. The writer and many others who have observed the phenomena discussed in this paper exercised these

precautions and experienced no discomfort except for afterimages that last for a few minutes.

SUMMARY

Evidence concerning the absorption of light quanta by visual pigment molecules is used to derive a theoretical relationship between the parameters of light incident on the retina and the output of the receptor cells. This relationship leads to the following prediction: when two retinal areas are illuminated, one very intensely and the other still more intensely, there will be a period of time during which the less intensely illuminated region will appear brighter than the more intensely illuminated one. Data verifying this prediction are presented.

When the theoretical treatment is expanded to include certain assumptions about retinal color mechanisms, it provides an explanation of some of the changes in hue that occur when a very intense colored stimulus is viewed steadily. Data on the temporal course of these hue changes are presented, and their relationship to the theory is discussed. This kind of analysis is shown to be a possible means for the quantitative evaluation of certain physiological properties of the visual system.

REFERENCES

AUERBACH, E., & WALD, G. The participation of different types of cones in human light and dark adaptation. *Amer. J. Ophthal.*, 1955, **39**, 24–40.

BRINDLEY, G. *Physiology of the retina and visual pathway.* London: Edward Arnold, 1960.

COHEN, J. Color adaptation to 1945. *Psychol. Bull.*, 1946, **43**, 121–140.

CORNSWEET, T. N., FOWLER, H., RABEDEAU, R. G., WHALEN, R. E., & WILLIAMS, D. R. Changes in the perceived color of very bright stimuli. *Science*, 1958, **128**, 898–899.

DEMOTT, D. W., & BOYNTON, R. M. Retinal distribution of entoptic stray light. *J. Opt. Soc. Amer.*, 1958, **48**, 13–21.

HECHT, S. A theory of visual intensity discrimination. *J. gen. Physiol.*, 1935, **18**, 767–789.

KROPF, A., & HUBBARD, R. The mechanism of bleaching rhodopsin. *Ann. N. Y. Acad. Sci.*, 1958, **74**, 266–280.

NODDACK, W., & JARZYK, G. Der Nachweis der Farbsehstoffe in lebenden menschlichen Auge. *Farbe*, 1955, **4**, 59–78.

RUSHTON, W. A. H. Kinetics of cone pigments measured objectively on the living human fovea. *Ann. N. Y. Acad. Sci.*, 1958, **74**, 291–304.

TROLAND, L. T. *Principles of psychophysiology.* Vol. 2. New York: Van Nostrand, 1930.

WHITESIDE, T. C. D. *Dazzle from nuclear weapons.* Washington, D. C.: National Academy of Science-National Research Council, 1960. (NAS-NRC Publ. No. 835)

(Early publication received October 27, 1961)

Psychological Review
1962, Vol. 69, No. 4, 274–294

PRACTICE AS A PROCESS OF SIMPLIFICATION

MARSHALL B. JONES[1]

United States Naval School of Aviation Medicine, Pensacola, Florida

From time to time in the last 30 years a psychologist has intercorrelated successive trials in the practice of some task or other. And in every instance the correlations have arranged themselves in a very definite pattern.

The correlations which appear in Table 1 are a good example. The task here is the Two-Hand Coordination Test, which requires the subject to move a pin around a clover shaped runway by the coordinated turning of two control handles. The subjects are 152 basic airmen; and the trials are 60 seconds long, with 30-second rest intervals between trials.[2]

TABLE 1

INTERTRIAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.77	.74	.73	.71	.71	.70
X_2			.87	.87	.84	.82	.82	.82
X_3				.91	.89	.87	.85	.86
X_4					.91	.88	.86	.88
X_5						.89	.90	.90
X_6							.93	.93
X_7								.94
X_8								

But notice the pattern. For each trial the correlations are highest in the superdiagonal (just above the main diagonal) and decrease monotonically as we move either upwards or to the right, until we come to the smallest correlation, which is in the upper right-hand corner of the matrix. There are a few inversions in the last two columns but they are minor, amounting in no case to more than .02.

This pattern, this "superdiagonal form" as I will call it, can be seen in every practice matrix which has ever been published. It seems to be a universal property of practice that a trial correlates strongest with its neighbors and weakest with the first and last trials in the series (Edgerton & Valentine, 1935; Fleishman, 1953, 1960; Fleishman & Hempel, 1954, 1955; Fleishman & Parker, 1959; Greene, 1943; Houston, 1950; Melton, 1947, p. 1022; Perl, 1934; Reynolds, 1952a, 1952b; Viteles, 1933).

PRINCIPAL POSSIBILITIES

Now what are we to make of this fact? Structurally, there are three principal possibilities. The first, in an historical sense, is the Woodrow two-factor structure.

In 1939 Herbert Woodrow observed that every practice matrix which was known at that time could be explained in terms of two common factors with positive weights which increased on one factor and decreased on the other as one went through the practice series. In other words, he posited that:

$$X_1 = a_1F_1 + b_1F_2 + c_1S_1$$
$$X_2 = a_2F_1 + b_2F_2 + c_2S_2$$
$$X_3 = a_3F_1 + b_3F_2 + c_3S_3$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$X_n = a_nF_1 + b_nF_2 + c_nS_n$$

[1] Opinions or conclusions contained in this paper are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

[2] For the use of these data I am indebted to E. A. Bilodeau.

where the X_i were the trial scores, F_1 and F_2 were the two common factors, the S_i were specific factors, and loadings, a_i and b_j, obeyed these laws:

$$0 \leq a_1 \leq a_2 \leq \cdots \leq a_{n-1} \leq a_n$$

and

$$b_1 \geq b_2 \geq b_3 \geq \cdots \geq b_n \geq 0$$

We now know that the Woodrow structure cannot explain all practice matrices. Sometimes, particularly in long matrices, a third Woodrow factor is required, or even a fourth. Nevertheless, the great bulk of practice matrices can be explained in terms of two Woodrow factors, and the present matrix is a case in point. If we hypothesize that:

$a_1 = .61$ $a_5 = .92$ $b_1 = .57$ $b_5 = .24$
$a_2 = .75$ $a_6 = .94$ $b_2 = .57$ $b_6 = .19$
$a_3 = .83$ $a_7 = .94$ $b_3 = .47$ $b_7 = .19$
$a_4 = .87$ $a_8 = .94$ $b_4 = .39$ $b_8 = .19$

we get the correlations which appear in Table 2. Clearly, the fit is good. So, as far as the correlations are concerned, the Woodrow structure is a tenable hypothesis.

The second principal possibility is the "double accumulation structure." For an eight-variable matrix the equations are:

$$X_1 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + s_1$$
$$X_2 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_{12} + s_2$$
$$X_3 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_{12} + f_{13} + s_3$$
$$X_4 = f_1 + f_2 + f_3 + f_4 + f_5 + f_{11} + f_{12} + f_{13} + s_4$$
$$X_5 = f_1 + f_2 + f_3 + f_4 + f_{10} + f_{11} + f_{12} + f_{13} + s_5$$
$$X_6 = f_1 + f_2 + f_3 + f_9 + f_{10} + f_{11} + f_{12} + f_{13} + s_6$$
$$X_7 = f_1 + f_2 + f_8 + f_9 + f_{10} + f_{11} + f_{12} + f_{13} + s_7$$
$$X_8 = f_1 + f_8 + f_9 + f_{10} + f_{11} + f_{12} + f_{13} + s_8$$

where the trial scores are in standard form and all factor scores are in *deviation* form only.[3] In some respects double accumulation resembles Woodrow structure. Both hypotheses call for a two-process interpretation of practice, in which one process gains in importance with practice and the other loses. They differ, however, in the way these two processes are conceived. In the Woodrow structure each process is represented by a single factor with weights which decrease or increase as practice proceeds. In the double accumulation structure each process is represented by a set of factors in which the included factors accumulate or drop out with each new trial.

The double accumulation structure is a special case of the "open contiguity structure," and there are some practice matrices which require the more general form. In the present example, however, if we hypothesize that $\sigma_{f_1}^2 = .70$ and that:

$\sigma_{f_2}^2 = .00$ $\sigma_{f_5}^2 = .02$ $\sigma_{f_8}^2 = .01$ $\sigma_{f_{11}}^2 = .02$
$\sigma_{f_3}^2 = .00$ $\sigma_{f_6}^2 = .02$ $\sigma_{f_9}^2 = .03$ $\sigma_{f_{12}}^2 = .04$
$\sigma_{f_4}^2 = .03$ $\sigma_{f_7}^2 = .02$ $\sigma_{f_{10}}^2 = .02$ $\sigma_{f_{13}}^2 = .12$

TABLE 2

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A WOODROW HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.78	.77	.75	.70	.68	.68	.68
X_2			.89	.87	.83	.81	.81	.81
X_3				.91	.88	.87	.87	.87
X_4					.89	.89	.89	.89
X_5						.91	.91	.91
X_6							.92	.92
X_7								.92
X_8								

[3] In the double accumulation structure and the simplex the right-hand side of each structural equation should be multiplied by an arbitrary constant. Otherwise, the structure conditions the variances and covariances as well as the correlations. Ordinarily, these constants are dropped as their inclusion would only complicate the equations to no particular advantage. Whenever we deal with variances or covariances, however, these constants must be reinstated.

TABLE 3

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A DOUBLE-ACCUMULATION HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.77	.75	.73	.70	.70	.70
X_2			.89	.87	.85	.82	.82	.82
X_3				.91	.89	.86	.86	.86
X_4					.91	.88	.88	.88
X_5						.90	.90	.90
X_6							.93	.93
X_7								.94
X_8								

the double accumulation structure generates the correlations which appear in Table 3. In this instance, therefore, as in most practice matrices, the correlations are easily explained in terms of a double accumulation hypothesis.

The third principal possibility, the simplex, differs fundamentally from the first two in that it calls for a one-process interpretation of practice. Practice, under a simplicial interpretation, is a process of simplification, exclusively. For an eight-variable matrix the equations are:

$$x_1 = f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 + s_1$$
$$x_2 = f_2 + f_3 + f_4 + f_5 + f_6 + f_7 + f_8 + s_2$$
$$x_3 = f_3 + f_4 + f_5 + f_6 + f_7 + f_8 + s_3$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$x_7 = f_7 + f_8 + s_7$$
$$x_8 = f_8 + s_8$$

where *all* variables are in deviation form only.[4]

[4] The requirement in the simplex that all variables be in deviation form only is important. Were we to require that the trial scores be expressed in standard form or, which is the same thing, that the trial scores all have the same variance (except for the arbitrary constants), then the "simplex" would become a single accumulation structure —which is a perfectly valid structure—but it has no place in the context of practice because it does not generate a superdiagonal form. For more on the single accumulation structure, see Jones (1960, pp. 50–51).

To generate specific correlations from a simplex we must hypothesize two kinds of parameters: the communalities

$$h_i^2 = \frac{\sigma_{f_i}^2 + \cdots + \sigma_{f_n}^2}{\sigma_{x_i}^2}$$

and the complexities

$$c_i = \left(\frac{\sigma_{f_i}^2 + \cdots + \sigma_{f_n}^2}{\sigma_{f_i}^2 + \cdots + \sigma_{f_n}^2}\right)^{\frac{1}{2}}$$

In the present instance, if we set:

$h_1^2 = .79$	$h_5^2 = .91$	$c_1 = 1.00$	$c_5 = .86$
$h_2^2 = .87$	$h_6^2 = .93$	$c_2 = .93$	$c_6 = .83$
$h_3^2 = .91$	$h_7^2 = .94$	$c_3 = .90$	$c_7 = .83$
$h_4^2 = .91$	$h_8^2 = .94$	$c_4 = .88$	$c_8 = .83$

we get the correlations which appear in Table 4 and a good fit. This particular example, moreover, is not at all exceptional. Every known matrix of correlations between trials of practice can be fit simplicially.

I should also point out that, right now, there does not exist any way of testing the significance of these fits. However, no one of these hypotheses leaves a residual correlation greater than .02, and we hardly need a test to tell us that residuals of this magnitude are not significant. Even if they were significant, the fact would be of no particular moment. True, the hypotheses would have to be modified;

TABLE 4

THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST UNDER A SIMPLICIAL HYPOTHESIS

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.77	.77	.75	.73	.71	.71	.71
X_2			.86	.85	.83	.81	.81	.81
X_3				.89	.87	.86	.86	.86
X_4					.89	.87	.87	.87
X_5						.89	.89	.89
X_6							.93	.93
X_7								.94
X_8								

but the modifications would be so paltry in amount that nothing of any real consequence would be altered.

MOLAR CORRELATIONAL ANALYSIS

To the learning theorist the next question is obvious: Which of these three structures best expresses the true nature of practice? I intend to answer this question, but first I want to say a word or two about correlational analysis.

For the last 30 years the treatment of correlation matrices in psychology has been completely dominated by factor analysis. There has been criticism, of course. In fact, there are a great many psychologists, perhaps even a majority, who won't have any part of factor analysis. But for two generations now there has been no alternative to factor analysis. If a man calculated a correlation matrix, the only existing way of summarizing his result in structural terms has been the factor analytic way. In the last few years, however, this situation has changed. Beginning with Louis Guttman's work on the simplex in 1954 and continuing with my two monographs on correlational analysis (Jones, 1959, 1960), an entirely new discipline has been created, a completely different way of treating a correlation matrix.

This new discipline, molar correlational analysis, is coextensive with factor analysis in practice and absolutely incompatible with it in theory. The essence of this incompatibility is a matter of approach. Factor analysis is an inductive method. The factor analyst starts with the correlations and then attempts by a perfectly mechanical computing regime to grind out his factor structure. He knows very well, of course, that there are a great many, substantially different structures any one of which could generate the correlations. But the factor analyst has faith. Aside from their intercorrelations he may know nothing whatever about the variables which enter into the matrix. Nevertheless, he remains quite confident that by estimating communalities, factoring the matrix, and rotating axes (according to his preferred method) he will arrive at the structure which is actually responsible for the correlations being what they are.

The molar analyst is a much less believing creature. As he sees it, correlational analysis is a strictly deductive discipline. If we are given some hypothesis as to how the variables are structured, then algebra alone can tell us what the correlations ought to be—but that's all it can do. The construction of theories is a business for the psychologist, as psychologist, not the statistician. All the statistician can do is to give us the means for tracing out the consequences of our thinking. He can tell us which patterns follow from which structures. He can catalogue the more important of these structures and these patterns. He can even show us how to fit a structure to a correlation matrix. But beyond this he simply cannot go.

In itself, this difference would enable us to construct an entirely different approach to the treatment of correlation matrices. But there are several other respects in which molar analysis differs fundamentally from the factor analytic tradition. The most important of these is the emphasis which the molar analyst places upon correlation pattern. A correlation matrix is more than a collection of correlations. As often as not, these correlations, or a substantial number of them, display a definite pattern. We have already seen the superdi-

agonal form. But there are many others as well: the monotonic hierarchy, for example, or circumplicial form, or rippled patterns, or tiered patterns, or three-cornered patterns. The factor analyst pays no attention to these patterns. To him it makes no difference how the matrix is patterned or even that it is patterned; under all circumstances he factors the matrix and rotates axes according to his a priori convictions. To the molar analyst, however, these patterns are an essential part, indeed, the most essential part, of his empirical results.[5]

Another basic difference between molar and factor analysis has to do with the idea of structure. If his thinking requires it, the molar analyst can and does work with factor structures of the conventional type, as for example, in the Woodrow two-factor structure. But he isn't content with these structures. The factors of a conventional structure bear no organic relationship one to the other. They aren't even mutually exclusive; each variable has so much of this factor and so much of that one, and that's it. But factor structures can also be organized, just as correlations can be. Consider, for example, the double accumulation structure or the simplex. These structures are much more than a list of factors. They are organized about principles, accumulation in the one case and inclusion in the other, which have a meaning altogether independent of the individual factors which make them up. These same principles, however, also operate in psychological theory. Ideas like contiguity, inclusion, accumulation, or competition are constantly recurring themes of psychological thought. But if we are to use these ideas in thinking about correlational results, we must have factor structures which can express them.

Altogether, the idea of molar correlational analysis is best understood through an analogy with curve fitting. In curve fitting we are presented initially with a plot of points. Ordinarily, these points have a shape, however obscured it may be by error. This shape may be unimodal, bimodal, **S** shaped, **U** shaped, or what have you; but whatever it is, the first thing that the curve fitter does is to look at the plotted points. Then, knowing something about analytic geometry and the nature of the two variables which are being related, he leaps to an hypothesis. This plot, he says, can be generated by a parabola, or a straight line, or a power function. There is, of course, no guarantee that he is right. There are always an infinity of functional forms which can fit any set of plotted points. But the curve fitter picks out one of these as his personal notion of what the relationship actually is. And finally, he attempts a quantitative fit. The fact that a plot of points is roughly **U** shaped does not mean that it can be fit by a parabola. First, we must posit values for the parameters of the parabolic function and generate a theoretical locus of points. Then, if the fit is good, we can conclude that the equation hypothesized by the worker is at least a tenable hypothesis.

Molar correlational analysis follows exactly the same sequence of events.

[5] When a correlation matrix comes off the machines, the variables which enter into it may or may not be ordered so as to reveal whatever patterns it contains. In a simplex, for example, the variables must be ordered from simple to complex or vice versa if the superdiagonal form is to appear. If the variables aren't ordered this way at first, then we must reorder them until we find the pattern. In my "Molar Correlational Analysis" (Jones, 1960, p. 75) I describe a device which makes it much easier to rummage around in a correlation matrix this way.

The correlations correspond to the plotted points; the pattern of the correlations corresponds to the shape of the plotted points; and the molar structure corresponds to the functional equations of the curve fitter. Like the curve fitter, the molar analyst begins by looking at his data. Then, knowing what the variables are that are being correlated and knowing something about which structures generate which patterns, he leaps to an hypothesis. This matrix is generated by a Woodrow structure, by a double accumulation structure, or by a simplex. Then finally, he attempts a quantitative fit. It he can posit values for the structural parameters which yield a good fit to the empirical correlations, then his hypothesis is one possible interpretation of the facts. If not, he can safely afford to reject it.

From a technical point of view, the central act of molar correlational analysis is fitting a structure to the correlation matrix. And let me emphasize once again that in this fitting the molar analyst is not looking for a structural hypothesis, as the factor analyst does when he factors the matrix and rotates axes. The molar analyst has already hypothesized his structure; his sole purpose is to determine the best fitting values for the structural parameters. In the body of this paper I will not present fitting procedures for any of the structures we have under consideration. However, I have appended a Supplementary Note in which these procedures are taken up at length.

Finally, let me put in a word about "simplicity." When we hypothesize a Woodrow structure, a double accumulation structure, or a simplex, we are speaking of the form of our hypothesis, in exactly the same sense that we speak of parabolic, linear, or logarithmic forms in curve fitting. The simplicity of the hypothesis is determined by the number of data we have to explain and the number of arbitrary structural parameters we must hypothesize to explain them. In the Woodrow structure there are $(2n)$ of these parameters, two factors with n loadings (the arbitrary parameters) on each; in the double accumulation structure there are $(2n-3)$ structural parameters, which in this case are the factor variances; and in the simplex there are $(2n-1)$ parameters,[6] the n communalities and the $(n-1)$ complexities (one of the complexities is arbitrary). All told, therefore, the double accumulation structure is the simplest explanation; then comes the simplex, and lastly the Woodrow two-factor structure.

Let me emphasize, however, that in molar analysis this whole question of simplicity is a purely *technical* detail. The molar analyst, unlike the factor analyst, lays no great philosophic stress upon simplicity. For him, the task which confronts the correlational theorist is to propose true theories, not simple ones. His problem is not to explain the correlations in this one matrix in the simplest possible terms but, rather, to integrate the correlations with what else he knows about the correlated variables. And according as he succeeds or fails in this direction, he expects his theories to stand or fall.

[6] In the past I have not regarded the communalities in a simplex as structural parameters altogether on a par with the complexities. My presentation in this paper, therefore, marks a definite change of position. The principal consequence is that a simplex, under this new position is not as simple a structure as I had previously made it out to be, not by half. It was this consequence, I suppose,—and a lingering whisp of the factor analytic miasma—which led me into my original error, but an error it incontestably was.

SIMPLICIAL ARGUMENT

Let's return now to the problem at hand, the superdiagonal forms of practice and their structural interpretation. At the present stage of the game the arguments for and against our three main hypotheses are far from decisive. In my opinion, the simplex has the best of it on the available evidence, but we have a long way to go before any one of these hypotheses or, for that matter, any other explanation of the correlations can be regarded as definitely established.

The case for the simplex rests on two main arguments. The first, and most weighty, comes from experimental psychology. Since the days of Bekhterev, psychologists have known that learning involved the progressive focussing of behavior on the one response which was successful in securing reward. At first the dog starts wildly in the stock. He makes all sorts of inappropriate responses; indeed, initially, it is only by accident that he lifts his paw. As practice continues, however, his behavior gradually focalizes, until in the end the dog barely lifts his paw the amount required to avoid the shock. As we observe it, in other words, practice is a process of simplification. How very natural therefore, that in its differential composition (which we do not observe) practice should also become simpler and simpler as one trial follows another.

In making this argument I want to be quite clear that I am talking about practice, not learning. There are some forms of learning, problem solving for example, which almost certainly are not processes of simplification. If Krechevsky (1932) had intercorrelated successive trials in his famous studies of "hypothesis" formation in rats, it is quite possible that the correlations would have shown the superdiagonal form. But in this case the originating structure could hardly have been a simplex. The rats' "hypotheses," after all, came and went. First they tried one thing and then another, first a spatial hypothesis and then a visual one. This is the kind of behavior which generates an open contiguum, a structure in which factors can come as well as go, not a simplex, in which they can only go (or come).[7] By "practice" I mean that portion of the learning sequence which follows the formation of the correct hypothesis. Learning doesn't end just because the person or animal has finally understood the problem. There remains the "stamping-in," the mastery of the problem. And it is this stamping-in, this practice, which the simplicial hypothesis requires to be a process of simplification.

The second argument has to do with the external relations of early and late stages of practice. According to the simplicial hypothesis, the early stages of practice contain a variety of factors in addition to those which are most particularly required in the practice task itself. This hypothesis implies that, if we correlate the trials of practice with materials which are categorically distinct from the practice task, then, on the average, the early stages of practice should be more strongly related to the test materials than the later ones. On the other hand, if we

[7] If a correlation matrix can be explained by a simplex with progressive simplification, it can also, as a matter of analytic necessity, be explained by a simplex with progressive complications. From a correlational point of view alone the possibility cannot be dismissed that the trials become increasingly complex with practice. I have rejected this notion for the same reasons I have accepted its inverse of progressive simplification, namely, its contradiction by the overt behavior of the subjects and by the correlations of the trials of practice with external measures.

correlate the trials of practice with very similar materials, we should expect just the reverse result: in general, the later stages should be more strongly related to the test materials than the early ones.

Unfortunately, the only evidence we have which is relevant to these two consequences concerns psychomotor tasks. In this one area, however, the evidence is very clear-cut. When they used printed tests as the external measures, Adams (1953, 1957), Fleishman (1960), Fleishman and Hempel (1954, 1955), and Reynolds (1952a, 1952b) all found that the predictability of the trials on a psychomotor task generally decreased with practice. On the other hand, these same investigators have also shown that the relationship between one psychomotor task and another usually increases with practice, on either one of the two tasks or both. And if we take an area more or less in between, say training criteria, we find a more or less neutral result: for example, the correlation between the Air Force pass-fail pilot criterion and scores on the Complex Coordination Test neither increases or decreases with practice. Unless we make special assumptions, however, none of these results follows from either the Woodrow or the double accumulation structure. All told, therefore, the existing evidence on the relationship between trials of practice and external measures strongly supports the simplicial hypothesis.

IDENTIFYING FACTORS

In all branches of study the recognition of form is followed by the analysis into parts. Once the astronomer has recognized a star he wants to know what it's made of. Once the biochemist has isolated a substance he wants to know it's composition. And once the psychologist has decided, even tentatively, that practice is a process of simplification he wants to know what the elements are which participate in this process; he wants to identify the simplicial factors.

Now, the only way the psychologist can identify these f_i is to administer a variety of other tests to the same subjects who worked on the practice task. Once this is done, he can correlate the reference tests with the stages of practice and, through these correlations, obtain the correlations between the reference tests and the simplicial factors. If these final correlations are high, then the factors are identified. If not, we must try again, with other tests.

To illustrate this logic let us rework the results of a recent study by E. A. Fleishman.[8] In Table 5 I have reproduced the intertrial correlations on

[8] In his analysis of these data Fleishman threw both the trials of practice and the reference tests into one grand matrix and performed a factor analysis. There is, I think, a fatal objection to this procedure, quite apart from Fleishman's use of factor analysis. It is that the structure of the *practice task* becomes dependent upon the test materials. If we use one battery of tests, we will get one structure for the trials of practice; if we use another, we will get a second; if we use a third battery, we will get still another. But it's obvious, it seems to me, that the trials of practice on, say, the Rotary Pursuit Test have *some* structure, altogether apart from the tests we happen to administer. This structure doesn't have to be a simplex. As a factor analyst, Fleishman might very well prefer the Woodrow structure. But however we interpret them, it seems quite clear that there is only one true structure for a single set of variables. Under Fleishman's approach the factors we are trying to identify are a function of the materials with which we are trying to identify them. The only way to resolve this dilemma is the molar way, to analyze the intertrial correlations first, apart from the reference tests, and then, having arrived at a structural recognition of the factors, to attempt an indentification in substance.

TABLE 5

INTERTRIAL CORRELATIONS FOR THE ROTARY PURSUIT TEST

Trial[a]	1	3	5	7	9	11	13	15
1		.82	.73	.72	.65	.65	.58	.54
2			.88	.84	.77	.74	.70	.64
5				.90	.86	.83	.79	.74
7					.89	.86	.83	.80
9						.88	.87	.82
11							.88	.84
13								.85
15								

[a] Fleishman reported his results for the odd-numbered trials only.

TABLE 6

THEORETICAL CORRELATIONS FOR THE ROTARY PURSUIT TEST UNDER A SIMPLICIAL HYPOTHESIS

Trial	1	3	5	7	9	11	13	15
1		.79	.74	.72	.66	.63	.61	.56
3			.86	.83	.77	.73	.71	.65
5				.91	.85	.80	.77	.72
7					.91	.86	.83	.77
9						.91	.89	.82
11							.91	.84
13								.84
15								

the practice task, which was the Rotary Pursuit Test. This matrix has the superdiagonal form. And if we hypothesize a simplex and set:

$h_1^2 = .86$ $h_5^2 = .96$ $c_1 = 1.00$ $c_5 = .73$
$h_2^2 = .92$ $h_6^2 = .94$ $c_2 = .89$ $c_6 = .70$
$h_3^2 = .94$ $h_7^2 = .90$ $c_3 = .82$ $c_7 = .69$
$h_4^2 = .98$ $h_8^2 = .82$ $c_4 = .78$ $c_8 = .67$

we get the theoretical correlations which appear in Table 6. Since the fit is good, our hypothesis is tenable. It remains to find out what the hypothesized factors are, not structurally (we already know that) but substantially.

In addition to the Rotary Pursuit Test Fleishman administered a battery of 17 reference tests, 7 paper-and-pencil tests, and 10 apparatus tests.[9] Now,

[9] The reader will notice that for the printed tests, Tests 1 through 7, the early trials are more strongly related than the late ones,

TABLE 7

CORRELATIONS BETWEEN THE REFERENCE TESTS AND THE TRIALS OF PRACTICE

Test	Trials							
	1	3	5	7	9	11	13	15
General mechanics	.19	.22	.22	.18	.20	.19	.15	.12
Tool functions	.19	.22	.18	.16	.17	.14	.13	.07
Speed of identification	.20	.24	.26	.26	.29	.26	.27	.26
Instrument comprehension	.20	.22	.22	.20	.21	.25	.20	.23
Visual pursuit	.30	.33	.25	.23	.23	.24	.22	.21
Aiming	.27	.32	.34	.29	.32	.30	.27	.29
Marking accuracy	.26	.33	.33	.31	.34	.35	.38	.34
Purdue assembly board	.20	.21	.18	.16	.17	.15	.17	.17
Santa Ana dexterity	.28	.36	.39	.38	.32	.36	.37	.34
Rate control	.25	.25	.19	.21	.18	.17	.12	.19
Linear pursuitmeter	.23	.22	.18	.21	.20	.17	.18	.18
Rate of movement	.19	.22	.23	.21	.23	.21	.18	.26
Rotary aiming	.27	.25	.29	.27	.29	.26	.29	.29
Reaction Time I	.24	.24	.26	.24	.22	.19	.18	.25
Track tracing	.34	.34	.35	.34	.32	.32	.32	.33
Complex coordination	.37	.36	.34	.35	.32	.31	.35	.39
Reaction Time II	.21	.29	.32	.30	.27	.28	.32	.34

TABLE 8

CORRELATIONS BETWEEN THE REFERENCE TESTS AND THE SIMPLICIAL FACTORS

Test	Factors							
	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
General mechanics	.02	−.03	.00	.06	−.11	.00	.11	.21
Tool functions	−.07	.08	−.10	.03	−.04	−.21	.11	.21
Speed of identification	.05	.08	.00	.12	.00	.11	.21	.22
Instrument comprehension	.14	−.11	.20	.06	−.07	.00	.11	.22
Visual pursuit	.07	−.03	.07	.06	−.04	−.42	.11	.33
Aiming	.14	−.03	.03	.12	−.15	.11	.21	.30
Marking accuracy	.02	.19	.13	1.2	−.04	.00	.32	.29
Purdue assembly	.02	.05	.00	.06	−.07	−.16	.05	.22
Santa Ana dexterity	.02	.14	.23	−.12	.00	.16	.37	.31
Rate control	.19	−.05	.00	−.06	.11	−.32	.00	.27
Linear pursuitmeter	.05	.05	−.07	.06	.11	−.21	−.05	.25
Rate of movement	.28	.03	−.03	.06	−.04	.05	.11	.21
Rotary aiming	.09	.11	.00	.09	−.04	.21	−.16	.30
Reaction Time I	.26	.03	−.07	−.03	−.04	.11	.00	.26
Track tracing	.12	.05	.07	.00	.00	.05	.00	.37
Complex coordination	.21	.22	.00	−.03	.07	−.11	−.11	.41
Reaction Time II	.16	.19	.13	−.06	−.07	.16	.42	.23

if we assume that the test-trial correlations, which appear in Table 7, are mediated by the simplicial factors only, we can calculate the correlations between the tests and the simplicial factors (Jones, 1960, pp. 28–31, 80–83). These last correlations, of course, are not necessarily exact, because the assumption under which they were calculated could be in error. However, when the simplicial communalities are as high as they are for the Rotary Pursuit Test, ranging from .86 to .98—and the test-trial correlations are as low as they are in this instance—we run little risk of serious error in accepting the hypothesis of exclusive simplicial mediation. And under this hypothesis, the correlations between the reference tests and the simplicial factors take on the values which appear in Table 8.

These values, unhappily, are low. Not one of them ranges beyond the .40s and only three are that high. In this particular instance, therefore, our effort to identify the simplicial factors was a failure. Nevertheless, there is nothing the matter with the method; if we persist with it, then sooner or later we will find tests which correlate strongly with the simplicial factors. When this happens, we will have identified the f_i.

CONCLUSION

In the end, of course, hypothesizing structures and identifying their factorial elements will not suffice. If we really want to understand the superdiagonal forms of practice, there is nothing for it but to experiment with them. How would these patterns change as a function of distribution of practice, or of intermittent reward? How would they be affected by prior practice on a related task? Sooner or later we must augment the passive

whereas for the apparatus tests, Tests 8 through 17, just the reverse is true. In this case, these tendencies are weaker than they are in any other published study, but even so they are still there.

arts of correlational analysis with active experimentation. We must theorize, of course; and this theorizing will inevitably be correlational in character. But we must also experiment.

REFERENCES

ADAMS, J. A. The prediction of performance at advanced stages of training on a complex psychomotor task. *USAF Hum. Resour. Res. Cent. Res. Bull.*, 1953, No. 53-49.

ADAMS, J. A. The relationship between certain measures of ability and the acquisition of a psychomotor criterion response. *J. gen. Psychol.*, 1957, **56**, 121-134.

EDGERTON, H. A., & VALENTINE, W. L. A. A factor analysis of learning data. *Psychol. Bull.*, 1935, **32**, 719.

FLEISCHMAN, E. A. A factor analysis of intra-task performance on two psychomotor tests. *Psychometrika*, 1953, **18**, 45-55.

FLEISHMAN, E. A. Abilities at different stages of practice in rotary pursuit performance. *J. exp. Psychol.*, 1960, **60**, 162-171.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor task as a function of practice. *Psychometrika*, 1954, **19**, 239-252.

FLEISHMAN, E. A., & HEMPEL, W. E., JR. The relation between abilities and improvement with practice in a visual discrimination reaction task. *J. exp. Psychol.*, 1955, **49**, 301-312.

FLEISHMAN, E. A., & PARKER, J. F., JR. Prediction of advanced levels of proficiency in a complex tracking task. *USAF WADC tech. Rep.*, 1959, No. 59-255.

GREENE, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1943, **8**, 37-53.

GUTTMAN, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld, (Ed.), *Mathematical thinking in the social sciences*. Glencoe, Ill.: Free Press, 1954.

HOUSTON, R. C. An evaluation of the predictive properties of measures of variability of performance on three psychomotor tasks. Unpublished PhD thesis, University of Maryland, 1950.

JONES, M. B. Simplex theory. *USN Sch. Aviat. Med. Monogr.*, 1959, No. 3.

JONES, M. B. Molar correlational analysis. *USN Sch. Aviat. Med. Monogr.*, 1960, No. 4.

KRECHEVSKY, I. "Hypotheses" vs. "chance" in the pre-solution period in sensory discrimination. *U. Calif. Publ. Psychol.*, 1932, **6**, 27-44.

MELTON, A. W. (Ed.) *Apparatus tests*. Washington: United States Government Printing Office, 1947.

PERL, R. E. An application of Thurstone's method of factor analysis to practice series. *J. gen. Psychol.*, 1934, **11**, 209-212.

REYNOLDS, B. Correlations between two psychomotor tasks as a function of distribution of practice on the first. *J. exp. Psychol.*, 1952, **43**, 314-348. (a)

REYNOLDS, B. The effect of learning of the predictability of psychomotor performance. *J. exp. Psychol.*, 1952, **44**, 189-198. (b)

VITELES, M. S. The influence of training on motor test performance. *J. exp. Psychol.*, 1933, **16**, 556-564.

(Received October 19, 1960)

SUPPLEMENTARY NOTE

In this note I propose to take up the fitting procedures for the three structures which were treated in the text of this paper, the Woodrow two-factor structure, the double accumulation structure, and the simplex. As a necessary preliminary, I will also treat a few, more elementary forms.

Mathematically, the best way to fit any molar structure is by the method of least squares. Given a structural hypothesis, we know that the theoretical correlations, ρ_{ij}, are some function of the structural parameters, $p_1, p_2, \cdots, p_n$. We could, therefore, proceed to minimize the squares of the differences between the theoretical and the empirical correlations, i.e., between the ρ_{ij} and the r_{ij}. To do this we need only differentiate the expression:

$$\sum_i \sum_j (\rho_{ij} - r_{ij})^2$$

and solve the normal equations which result.

There is, however, a difficulty with this approach: The number of normal equations which must be solved in order to determine a best fit is prohibitive. If we use least squares, every fitting process

requires us to solve at least as many simultaneous equations as there are variables in the matrix. And this, in its turn, requires computing equipment of a high order. It is desirable, therefore, to find some other, less heroic means of arriving at essentially the same results.

In all the fitting procedures which I will describe the general process is the same. We begin by positing a tentative set of parameters, which we then adjust until we get what appears to be a best fit. These tentative values, of course, are not taken at random. Knowing what the structure is, we can determine tentative values which will require a minimum of adjustment. The only problem is to develop sound procedures for arriving at these tentative values.

SPEARMAN UNIT FACTOR

Let us begin with the Spearman unit factor, i.e., with the hypothesis that the variables which enter into the matrix contain a single common factor. If the matrix is arranged so that the variable with the biggest loading on this one factor comes first, the variable with the next biggest loading second, and so on, down to the variable with the smallest loading, which comes last, the structure generates a characteristic pattern. This pattern, which I call a "monotonic hierarchy," locates the biggest correlation in the upper left-hand corner of the matrix and the smallest in the lower right-hand corner with the others spread out in between. If the loadings are arranged the other way around, i.e., with the smallest loading first and the biggest last, then the pattern will be reversed: The biggest correlation will be in the lower right-hand corner and the smallest in the upper left-hand corner.

TABLE A1

A MONOTONIC HIERARCHY

Trial	X_1	X_2	X_3	X_4
X_1		.63	.64	.66
X_2			.70	.72
X_3				.74
X_4				

TABLE A2

SOME THEORETICAL CORRELATIONS UNDER A SPEARMAN UNIT-FACTOR INTERPRETATION

Trial	X_1	X_2	X_3	X_4
X_1		.63	.64	.66
X_2			.70	.72
X_3				.73
X_4				

The matrix in Table A1 displays this second form of the monotonic hierarchy. On a purely qualitative basis, therefore, the hypothesis of a Spearman unit factor is tenable. But before we can accept this hypothesis as a possible explanation of the correlations, we must first specify values for the factor loadings and attempt a fit.

Under a Spearman hypothesis the empirical correlations:

$$r_{ij} \approx a_i a_j$$

where a_i is the loading on the one common factor of the ith variable in the matrix. It follows that:

$$a_1 \approx \left(\frac{r_{12} r_{13}}{r_{23}} \right)^{\frac{1}{2}}$$

$$a_1 \approx \left(\frac{r_{1(n-1)} r_{1n}}{r_{(n-1)n}} \right)^{\frac{1}{2}}$$

In other words, working with the two "ends" of the matrix we obtain two tentative values for the first factor loading, a_1. If we compromise these two values, we will get a "balanced" single value, which, in this instance, equals .76. Then, since:

$$a_j \approx \frac{r_{ij}}{a_1}$$

we can go ahead and determine the tentative values for the remaining variables. In the case at hand we get these values:

$$a_1 = .76 \qquad a_3 = .84$$

$$a_2 = .83 \qquad a_4 = .87$$

Generating the theoretical correlations from these parameters we get the matrix

which appears in Table A2. Ordinarily, we are able to improve upon the fit by adjusting the a_i, but in this instance the fit is just about perfect as it stands, since the biggest residual correlation is .01, and only one is that big. It's quite clear, therefore, that the matrix can be explained as originating in a single common factor.

AUGMENTED SPEARMAN UNIT FACTOR

By an augmented Spearman unit factor I mean a two-factor structure in which the loadings of all the variables on one of the two factors are equal. Correlationally, the effect of this second factor is to add a constant quantity, which equals the loading on this second factor squared, to all of the correlations in the matrix. Now, consider the matrix which appears in Table A3. Qualitatively, the pattern it displays, i.e., the monotonic hierarchy, is consistent with an augmented-Spearman interpretation. It remains to be seen, however, whether or not a good quantitative fit is possible.

In fitting an augmented Spearman hypothesis our first task is to determine the magnitude of the one loading on the second factor. Consider, therefore, the three "principal points" of the matrix, i.e., the upper left-hand corner, the upper right-hand corner, and the lower right-hand corner. Under a *nonaugmented* Spearman hypothesis:

$$r_{12}r_{(n-1)n} - r_{1n}r_{2(n-1)} \approx 0$$

If this quantity is either negative or zero, then the fit cannot be improved by means of the additive constant (which must be positive). If this difference is positive, however, as in the present instance, then by subtracting a fixed positive amount from all the correlations in the matrix, the theoretical correlations can be brought more nearly into line with their empirical counterparts. Now, if we subtract .10 from all the correlations in the matrix, the tetrad difference between the principal points remains positive. In fact, it remains positive even if we subtract .50 from all of the correlations. If we subtract .60, however, the difference becomes negative. Suppose, therefore, that we take .50 as our tentative value for b^2, the one loading on the second factor squared. If we subtract this value from all of the correlations in the matrix and fit a Spearman unit factor to the residual correlations, in the manner already described, we get these values for our five parameters:

$$a_1 = .66 \quad a_3 = .41 \quad b = .71$$
$$a_2 = .58 \quad a_4 = .32$$

which generate the correlations which appear in Table A4. Once again, there is no need to adjust the tentative values for the parameters. But again it should be understood that adjustments are ordinarily required.

TABLE A3

ANOTHER MONOTONIC HIERACHY

Trial	X_1	X_2	X_3	X_4
X_1		.88	.77	.71
X_2			.73	.67
X_3				.63
X_4				

TABLE A4

SOME THEORETICAL CORRELATIONS UNDER AN AUGMENTED SPEARMAN UNIT-FACTOR INTERPRETATION

Trial	X_1	X_2	X_3	X_4
X_1		.88	.77	.71
X_2			.74	.69
X_3				.63
X_4				

HALF-WOODROW TWO-FACTOR STRUCTURE

In the half-Woodrow, two-factor structure the loadings on one of the two factors are the same for the first half of the series, while on the other factor the loadings are the same for the second half of the series. In an eight-variable structure, for example, the loadings on the

TABLE A5

A THREE-CORNERED PATTERN

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.88	.77	.71	.69	.70	.72	.74
X_2			.73	.67	.65	.67	.69	.70
X_3				.63	.62	.64	.65	.67
X_4					.60	.61	.63	.64
X_5						.63	.64	.66
X_6							.70	.72
X_7								.74
X_8								

first factor:

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$$

and the loadings on the second factor:

$$\beta_5 = \beta_6 = \beta_7 = \beta_8$$

For the rest of it:

$$\alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \alpha_8$$

and

$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$$

just as in the full-Woodrow structure.

Correlationally, however, the two structures are very different. The full-Woodrow generates a superdiagonal form, while the half-Woodrow generates a three-cornered pattern. Table A5 presents an eight-variable example. The correlations are high in the three corners of the matrix and drop off as one moves from any one of them toward the middle of the last column, the middle of the first row, or the middle of the superdiagonal.

A three-cornered pattern is naturally divided into three submatrices: the two apical triangles, which contain the correlations *within* the two halves of the series, and the central block, which contains the correlations *between* them. In the two apical triangles the correlations are generated by augmented Spearman hypotheses. So, we already know how to fit these correlations. In fact, we already have. The two apical triangles of Table A5 are the same two monotonic hierarchies which we treated in the first two sections of this note. Therefore, we already have tentative parameters for the half-Woodrow structure, namely:

$$\alpha = .71 \qquad \beta_1 = .66$$
$$\alpha_5 = .76 \qquad \beta_2 = .58$$
$$\alpha_6 = .83 \qquad \beta_3 = .41$$
$$\alpha_7 = .84 \qquad \beta_4 = .32$$
$$\alpha_8 = .87 \qquad \beta = .00$$

where α represents the first four loadings (all equal) on the first factor and β represents the last four loadings (again, all equal) on the second factor.

These loadings, of course, were derived exclusively from the two apical triangles. To this point we have taken no account of the correlations in the central block. To do so, however, we need only generate the block correlations from our preliminary loadings and see what sort of a fit we get. If the fit is good, our preliminary loadings become final. If it isn't, we must adjust the factor loadings so as to compromise the errors which they generate in the apical traingles and in the central block. In the case at hand the fit is not good. But the remedy is easy. All we need do is assign β a somewhat higher value than zero. If we let β equal .20 and keep all the other loadings just what they were, we get the theoretical correlations which appear in Table A6, and a good fit. Of course, it might not have been possible to fit the central block and the apical triangles with a single set of loadings. If this had been the case, we would have had to reject the half-Woodrow hypothesis.

TABLE A6

SOME THEORETICAL CORRELATIONS GENERATED BY A HALF-WOODROW TWO-FACTOR STRUCTURE

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.88	.77	.71	.67	.72	.73	.75
X_2			.74	.69	.66	.71	.71	.73
X_3				.63	.62	.67	.68	.70
X_4					.60	.65	.66	.68
X_5						.67	.68	.70
X_6							.74	.76
X_7								.77
X_8								

FULL–WOODROW TWO–FACTOR STRUCTURE

Fitting a full-Woodrow structure depends upon one fact: If a correlation matrix is derived from a full-Woodrow structure, it is always possible to transform the correlations into a three-cornered pattern and to fit the transformed correlations with a half-Woodrow structure. Suppose, for example, that we multiply the loadings a_i and b_i in a full-Woodrow structure by k_i, where:

$$k_i = \frac{a_i}{a_i}$$

for the first $n/2$ variables and

$$k_2 = \frac{b_n}{b_i}$$

for the last $n/2$ variables. The result will be a half-Woodrow structure, with the loadings (in the eight-variable case):

$$\alpha_1 = a_1 \quad \alpha_5 = a_5\left(\frac{b_8}{b_5}\right) \quad \beta_1 = b_1 \quad \beta_5 = b_5$$

$$\alpha_2 = a_1 \quad \alpha_6 = a_6\left(\frac{b_8}{b_6}\right) \quad \beta_2 = b_2\left(\frac{a_1}{a_2}\right) \quad \beta_6 = b_5$$

$$\alpha_3 = a_1 \quad \alpha_7 = a_7\left(\frac{b_8}{b_7}\right) \quad \beta_3 = b_2\left(\frac{a_1}{a_3}\right) \quad \beta_7 = b_5$$

$$\alpha_4 = a_1 \quad \alpha_8 = a_8 \quad \beta_4 = b_4\left(\frac{a_1}{a_4}\right) \quad \beta_8 = b_5$$

And the correlations which this new structure generates, call them r_{ij}', will bear a simple relationship to the correlations of the original matrix, namely:

$$r_{ij}' = r_{ij}k_ik_j$$

It follows that to fit a full-Woodrow structure we must do three things: transform the superdiagonal form into a three-cornered pattern, fit the three-cornered pattern with a half-Woodrow, and then transform the loadings for the half-Woodrow back to the original full-Woodrow structure.

Our first task, then, is to find values for the k_i which are inversely proportional to the full-Woodrow loadings, a_1, a_2, a_3, a_4, and b_5, b_6, b_7, b_8. The word "proportional" is important, because it means that we have a great deal of latitude in transforming the superdiagonal form; more specifically, it means that the transformed correlations may be big, small, or middling, without any loss of rigor. As long as the new correlations can be fit by a half-Woodrow structure, whatever their absolute magnitude may be, then we can fit the original correlations with a full-Woodrow structure. But even so, how do we find a set of k_i which will do the trick?

Consider the correlations in the central block of a three-cornered pattern. They take this form:

	X_5	X_6	X_7	X_8
X_1	$(\alpha\alpha_5+\beta\beta_1)$	$(\alpha\alpha_6+\beta\beta_1)$	$(\alpha\alpha_7+\beta\beta_1)$	$(\alpha\alpha_8+\beta\beta_1)$
X_2	$(\alpha\alpha_5+\beta\beta_2)$	$(\alpha\alpha_6+\beta\beta_2)$	$(\alpha\alpha_7+\beta\beta_2)$	$(\alpha\alpha_8+\beta\beta_2)$
X_3	$(\alpha\alpha_5+\beta\beta_3)$	$(\alpha\alpha_6+\beta\beta_3)$	$(\alpha\alpha_7+\beta\beta_3)$	$(\alpha\alpha_8+\beta\beta_3)$
X_4	$(\alpha\alpha_5+\beta\beta_4)$	$(\alpha\alpha_6+\beta\beta_4)$	$(\alpha\alpha_7+\beta\beta_4)$	$(\alpha\alpha_8+\beta\beta_4)$

which the reader will immediately recognize as a perfectly additive linear pattern, as these terms are used in analysis of variance. What we seek, therefore, are values for the k_i which will transform the central block of the superdiagonal form into this pattern. Now, we know from the nature of the Woodrow structure that:

$$\alpha_5 \leq \alpha_6 \leq \alpha_7 \leq \alpha_8$$

and

$$\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$$

which means that the biggest r_{ij}' in the central block will be in the upper right-hand corner. The correlations will fall off as we go either to the left or down, until we come to the smallest r_{ij}' in the central block, which will be in the lower left-hand corner. We also know, however, that the correlations in the central block of a superdiagonal form have just the opposite pattern: The biggest correlation is in the lower left-hand corner and the smallest in the upper right-hand corner. It follows that the multiplicative constants, k_i, must obey these inequalities:

$$k_1 \geq k_2 \geq k_3 \geq k_4$$

and

$$k_5 \leq k_6 \leq k_7 \leq k_8$$

where the first four constants are multiplied only into the rows of the central block and the last four only into the columns. Knowing this much there is no great difficulty in transforming the correlations so that they have the prescribed pattern.

Take the correlations in Table A5, for example. By now it is quite obvious that the correlations in Table A5 were derived from the original correlations for the Two-Hand Coordination Test. To produce the correlations of Table A5 each correlation r_{ij} in Table 1 was multiplied by k_i and again by k_j where:

$k_1 = 1.17$	$k_5 = .81$
$k_2 = .95$	$k_6 = .86$
$k_3 = .86$	$k_7 = .88$
$k_4 = .81$	$k_8 = .90$

All we have to do now, therefore, is to divide the loadings of the half-Woodrow structure, the α_i and β_i, by their corresponding k_{ij} and we have our tentative parameters for the original full-Woodrow structure, viz.:

$a_1 = .61$	$a_5 = .94$	$b_1 = .56$	$b_5 = .25$
$a_2 = .75$	$a_6 = .97$	$b_2 = .61$	$b_6 = .23$
$a_3 = .83$	$a_7 = .95$	$b_3 = .48$	$b_7 = .23$
$a_4 = .88$	$a_8 = .97$	$b_4 = .40$	$b_8 = .22$

These values do not conform entirely to the requirements of the full-Woodrow structure. b_1 should not be less than b_2; nor should a_7 be less than a_6. Also, the fit, though good, can be improved. However, we can make adjustments for these things; and when we do, we get the loadings which are cited in the text and the correlations which are reproduced in Table 2.

DOUBLE ACCUMULATION STRUCTURE

If a correlation matrix is generated by a double accumulation structure, then:

$$r_{ij} \approx r_{1n} + (r_{ij} - r_{1n}) + (r_{in} - r_{1n})$$

If the fit is perfect, the relationship is exact. In other words, the correlations generated by a double accumulation structure have the same additive linear pattern which we remarked in the central block of the half-Woodrow pattern. This circumstance makes it possible for us to fit the correlations under a double accumulation hypothesis without ever specifying the parameters of the structure, i.e., the variances of the factors.

Let's take a few correlations from the matrix in Table 1 and see how well they satisfy this linear, additive relationship. We have:

$$r_{23} = .87 \approx .70 + (.07 + .12) = .89$$
$$r_{34} = .91 \approx .70 + (.04 + .06) = .90$$
$$r_{45} = .91 \approx .70 + (.03 + .18)$$
$$r_{37} = .85 \approx .70 + (.01 + .16) = .87$$

It's quite obvious from these relationships that Table 1 can be accounted for as a double accumulation structure. All we have to do is to treat the correlations in the first row and last column as theoretical correlations and adjust them until we get a good fit, which, in the present case, is hardly much of a chore. Once they are made, we have the theoretical correlations, ρ_{ij}, which appear in Table 3, of the text.

It remains, of course, to determine the variances of the factors. In a double accumulation structure the first and last variables in the matrix have only one factor in common and that one factor is common to all of the variables in the matrix. It follows that the variance of this one factor equals r_{1n}. Now, apart from this general factor, a double accumulation structure contains $2(n-2)$ other factors, which are divided into two accumulation structures, each one containing $(n-2)$ factors. It happens, however, that the theoretical correlations in the first row of a double accumulation pattern are determined by one of these structures only, and the theoretical correlations in the last column are determined by the other, exclusively. This means that we can reel off the factor variances in a very simple way, once we know the theoretical correlations. If the factors are numbered the way they are in

the text, we get:

$$\sigma_{f_2}^2 = \rho_{17} - \rho_{18}$$
$$\sigma_{f_3}^2 = \rho_{16} - \rho_{17}$$
$$\cdots\cdots\cdots\cdots$$
$$\sigma_{f_7}^2 = \rho_{12} - \rho_{13}$$

and

$$\sigma_{f_{13}}^2 = \rho_{28} - \rho_{18}$$
$$\sigma_{f_{12}}^2 = \rho_{38} - \rho_{28}$$
$$\cdots\cdots\cdots\cdots$$
$$\sigma_{f_8}^2 = \rho_{78} - \rho_{68}$$

From these formulae I derived the values of the $\sigma_{f_i}^2$ which appear in the text.

SIMPLEX

In a Woodrow structure the parameters are all of the same sort, namely factor loadings; similarly, in the double accumulation structure, the parameters are also of the same sort, in this case factor variances. But in the simplex there are two different kinds of structural parameters. Under a simplicial hypothesis, the theoretical correlations:

$$\rho_{ij} = \frac{c_j}{c_i}\sqrt{h_i^2 h_j^2} \qquad (i < j)$$

where the c_i are the complexities and the h_i^2 are the communalities. This parametric duality creates a corresponding duality in the fitting procedures: We can specify the communalities first or we can specify the complexities first. The older method is to specify the communalities first. So I will begin there.

Communality Method

Using the communality method we fit a simplex in the following manner. First, we specify tentative values for the communalities. Then, the correlations are corrected for attenuation (by the communalities) and the resulting, unattenuated correlations fit with a perfect simplicial form.[A1] If the unattenuated correlations won't submit to a simplicial fit, it may be that the variables which enter into the matrix are not structured simplicially. It may also be that we used the wrong communalities. So we adjust the communalities. If, by these adjustments, we can bring the unattenuated correlations into simplicial form or an approximation to it, we have our fit. If we can't, then the simplicial hypothesis must be rejected.

It's quite obvious from this account that a great deal hinges on the original choice of communalities. If the matrix has simplicial form, then sooner or later we will find a fit. But if the values we posit for the communalities are sufficiently far off, it may be much later. It would be nice if we had some way of knowing at the beginning what these communalities would turn out to be. Unfortunately, in advance of trying them out, the best we can do is to specify limits for the communalities. It would not be reasonable, for example, to set the communality of a variable higher than its known reliability. Nor would it be resonable to set it so low that the correlation between two variables, after correcting for the communalities, became greater than unity. But within these limits there is no way of knowing in advance what the best-fitting values for the communalities are.

Nevertheless, we can still posit values for the communalities and try them out. And this is just what we do. In practice, there are three basic ways of proceeding. One, we can set the communality of the ith variable equal to the highest correlation it has in the matrix. Two, we can equate the communalities with one kind or another of reliability (if we have any information on this count). Or three, we can set the communalities all equal to a given value, .90, say, or .80. No one of these procedures has any a priori claim to being right. But they do represent so many ways of approaching the problem, and in my experience they are good ways. I have fit a great many simplexes, and one or another of these approaches, with minor adjustments, has always led to a good fit, if one existed.

In fitting the matrix for the Two-Hand Coordination Test I set the communali-

[A1] The term "perfect simplicial form" is defined on the next page. For a fuller discussion of the simplex and its terminology, see Jones (1959, pp. 11–15).

ties of the trials equal to the largest correlation they had in the matrix. The next step was to correct every correlation in the matrix for its attenuation by the remaining, specific variance. In other words, I divided r_{ij} by $\sqrt{h_i h_j}$. The matrix of unattenuated coefficients which resulted appears in Table A7.

Now, once the correlations have been corrected for attenuation the equations for the simplex become:

$$\begin{aligned} x_1 &= f_1 + f_2 + \cdots + f_{n-1} + f_n \\ x_2 &= \quad f_2 + \cdots + f_{n-1} + f_n \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_{n-1} &= \quad\quad\quad\quad f_{n-1} + f_n \\ x_n &= \quad\quad\quad\quad\quad f_n \end{aligned}$$

or just what they were originally, except that the specific factors are gone. In this form, the simplex lays down very demanding requirements on the correlations. If the unattenuated correlations r_{ij}', really do form a simplex, then they must obey the fundamental law of "perfect simplicial form," i.e.:

$$r_{ik}' = r_{ij}' r_{jk}'$$

For example, if $r_{12}' = .90$ and $r_{23}' = .80$, then r_{13}' must equal .72. Clearly, if this is the case, the matrix will have the superdiagonal form—and then some.

Of course, in practice the unattenuated coefficients never do have simplicial form, not exactly. Take the correlations of Table A7, for example. r_{12}' times r_{23}' equals .93, which almost equals $r_{13}' =$ .91, but not exactly. Our task, therefore, is to find a theoretical simplicial form which comes as close as possible to the unattenuated pattern. In the case of Table A7 this task is easy. The correlations, which appear in Table A8 have perfect simplicial form, and, as the reader can see, they constitute a very close approximation to the unattenuated empirical correlations.

TABLE A7

UNATTENUATED EMPIRICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COMMUNALITY METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.95	.91	.87	.86	.83	.83	.81
X_2			.98	.98	.94	.91	.91	.91
X_3				1.00	.98	.95	.92	.93
X_4					1.00	.96	.93	.96
X_5						.97	.98	.98
X_6							1.00	1.00
X_7								1.00
X_8								

TABLE A8

UNATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COMMUNALITY METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.93	.90	.88	.86	.83	.83	.83
X_2			.97	.95	.93	.90	.90	.90
X_3				.98	.96	.93	.93	.93
X_4					.98	.95	.95	.95
X_5						.97	.97	.97
X_6							1.00	1.00
X_7								1.00
X_8								

To obtain the theoretical correlations, c_{ij}, which I presented in Table 4 of the text I simply multiplied the correlations in Table A8 by the same communalities which I used to unattenuate the empirical correlations.

But what about the complexities? Here we are, with the ρ_{ij} already derived, and we still haven't said anything about the complexities. The answer, of course, is that, without realizing it, perhaps, we have already posited the complexities. The complexity of the ith trial is nothing other than its theoretical correlation, *before attenuation*, with the first trial in the series; in other words, $c_j = c_{ij}$. In the case at hand, therefore, the complexities are the correlations in the first row of Table A8.[A2]

[A2] The virtue of the complexities is that, if we know them, then by the fundamental law of perfect simplicial form we know all of the c_{ij} and can, therefore, if we also know the communalities, generate all of the theoretical correlations ρ_{ij}. However, the complexities are not alone in possessing this property. The gradations, i.e., the correlations $c_{i,i+1}$, are also capable of generating the entire set

Covariance Method[A3]

The full equations for the simplex (with progressive simplification) are:

$$x_i = k_i(f_i + \cdots + f_n + s_i)$$

where the k_i are arbitrary constants. In the usual presentation of the simplex, as in the text of this paper, these constants are dropped, since they are of no consequence as long as we have to do with correlations only. The moment, however, that we concern ourselves with variances or with covariances we must dredge these little things up and be most meticulous about them.

Under a simplicial hypothesis, the covariances between the variables:

$$\text{cov}_{ij} = k_i k_j(\sigma_{f_j}^2 + \cdots + \sigma_{f_n}^2)$$

Suppose, now, that we multiply all of the entries in the ith row by $\frac{k_o}{k_i}$, where k_o is another arbitrary constant. The new entries:

$$\text{cov}_{ij}^1 = k_o k_j(\sigma_{f_j}^2 \cdots + \sigma_{f_n}^2$$

And if we multiply the columns by the same quantities, we get the entries:

$$\text{cov}_{ij}^{11} = k_o^2(\sigma_{f_j}^2 + \cdots + \sigma_{f_n}^2$$

From here we get the unattenuated, theoretical correlations c_{ij} immediately, because:

$$c_{ij}^2 = \frac{(\sigma_{f_j}^2 + \cdots + \sigma_{f_n}^2)}{(\sigma_{f_i}^2 + \cdots + \sigma_{f_n}^2)}$$

The only problem is to find the $\frac{k_o}{k_i}$.

TABLE A9

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		323	331	329	376	353	360	375
X_2			477	484	546	510	523	528
X_3				536	617	575	580	584
X_4					640	593	598	613
X_5						702	726	729
X_6							718	722
X_7								751
X_8								

This problem, happily, is easily solved. Once we have multiplied through the rows by the $\frac{k_o}{k_i}$, then all of the entries in any one column are equal. So suppose we set all of the entries in the last column equal to some arbitrary amount A, by multiplying the rows by:

$$\frac{A}{\text{cov}_{in}} = \frac{k_o}{k_i}\left(k_o = \frac{A}{k_n\sigma_{f_n}^2}\right)$$

If the simplicial hypothesis is correct, then all of the entries in any one of the other columns should be equal too, or close to it. If this is not the case, we can reject the simplicial hypothesis. If it is, however, we have the $\frac{k_o}{k_i}$ we sought.

In Table A9 the covariance matrix for the Two-Hand Coordination Test is presented. Setting all the entries in the last column equal to 300 (except for rounding errors), we get these values for our $\frac{k_o}{k_i}$.

$$\frac{k_o}{k_1} = .84 \quad \frac{k_o}{k_3} = .51 \quad \frac{k_o}{k_5} = .41 \quad \frac{k_o}{k_7} = .40$$

$$\frac{k_o}{k_o} = .57 \quad \frac{k_o}{k_4} = .49 \quad \frac{k_o}{k_6} = .42$$

Note that we get no value for $\frac{k_o}{k_8}$. When the rows are multiplied by these values, we get the matrix which appears in Table A10. And since the columns conform approximately to the simplicial requirements, we know that we are on the right

---

of c_{ij} and so are several other subsets of the c_{ij}. For a discussion of this point, see Jones (1959, p. 12).

[A3] The first person to mention to me the possibilities of the covariance matrix in fitting the simplex to a correlation matrix was Ardie Lubin. My original reaction was very negative and in the "Simplex Theory" (Jones, 1959, pp. 79–80) I had some rather hard things to say about his suggestion. I have no course now, therefore, but to apologize to him for my intemperance: he was right and I was wrong.

TABLE A10

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST AFTER MULTIPLYING THROUGH THE ROWS BY $\frac{k_o}{k_i}$

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		271	278	276	316	297	302	300
X_2			272	275	311	291	298	301
X_3				273	315	293	296	298
X_4					314	291	293	300
X_5						288	298	299
X_6							302	303
X_7								300
X_8								

track. Now, multiplying through the columns by these same values we get the matrix which appears in Table A11. And from this table, taking the average down the columns, we get:

$$k_o^2(\sigma_{f_2}^2 + \cdots + \sigma_{f_8}^2) = 154$$
$$k_o^2(\sigma_{f_3}^2 + \cdots + \sigma_{f_8}^2) = 141$$
$$k_o^2(\sigma_{f_4}^2 + \cdots + \sigma_{f_8}^2) = 135$$
$$k_o^2(\sigma_{f_5}^2 + \cdots + \sigma_{f_8}^2) = 129$$
$$k_o^2(\sigma_{f_6}^2 + \sigma_{f_7}^2 + \sigma_{f_8}^2) = 123$$
$$k_o^2(\sigma_{f_7}^2 + \sigma_{f_8}^2) = 119$$

From these figures we immediately obtain the matrix of theoretical, unattenuated correlation coefficients (minus the first row and the last column) which appears in Table A12. It would have made

TABLE A11

THE COVARIANCE MATRIX FOR THE TWO-HAND COORDINATION TEST AFTER MULTIPLYING THROUGH THE ROWS AND THE COLUMNS BY $\frac{k_o}{k_i}$

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		154	142	135	130	125	121	
X_2			139	135	128	122	119	
X_3				134	129	123	118	
X_4					129	122	117	
X_5						121	119	
X_6							121	
X_7								
X_8								

no difference, by the way, if we had equated the rows first and then multiplied through the columns. We would have ended up with the same values for the c_{ij}.

Let's not worry for the time being about the first row and the last column and proceed directly to the communalities. If a simplicial fit exists, we know that:

$$r_{ij} \approx c_{ij}\sqrt{h_i^2 h_j^2}$$

If, therefore, we divide the r_{ij} by our already obtained c_{ij}, the resulting matrix should be a Spearman unit hierarchy, where the loadings for the several variables equal $\sqrt{h_i^2}$. Carrying out this process we find that:

$$h_2^2 = .92 \quad h_4^2 = .92 \quad h_6^2 = .94$$
$$h_3^2 = .92 \quad h_5^2 = .94 \quad h_7^2 = .94$$

It only remains to fill things in for the first and last variables.

The correlations in the first row:

$$r_{1j} \approx c_{12}c_{2j}\sqrt{h_1^2 h_j^2}$$

or

$$r_{1j} \approx l_1 c_{2j}\sqrt{h_j^2}$$

if we set

$$l_1 = c_{12}\sqrt{h_1^2}$$

In other words, given what we already know, i.e., c_{2j} and $\sqrt{h_j^2}$, for $2 \leq j \leq n-1$, the empirical correlations r_{1j} are generated by a *single* parameter. In the same

TABLE A12

UNATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COVARIANCE METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1								
X_2			.96	.94	.92	.89	.88	
X_3				.98	.96	.93	.92	
X_4					.98	.95	.94	
X_5						.97	.96	
X_6							.98	
X_7								
X_8								

way, the empirical correlations in the last column:

$$r_{in} \approx l_2 c_{i1(n-1)} \sqrt{h_i^2}$$

where

$$l_n = c_{(n-1),n} \sqrt{h_n^2}$$

and finally

$$r_{1n} = l_1 l_n c_{27}$$

All we have to do, therefore, is to obtain tentative values for l_1 and l_n. The most natural way to do this is to set:

$$l_1 = \frac{r_{12}}{\sqrt{h_2^2}},$$

$$l_n = \frac{r_{(n-1)n}}{\sqrt{h_{(n-1)}^2}}$$

and then solve for l_1 and l_n. In the present case, we get:

$$l_1 = .82,$$
$$l_8 = .97$$

Altogether, then, we can pick any values we like for c_{12} and h_1^2, provided only that:

$$c_{12} \sqrt{h_1^2} = .82$$

And in the same way we can pick any values we like for c_{78} and h_8^2, provided only that:

$$c_{78} \sqrt{h_8^2} = .97$$

Suppose, therefore, that we set:

$$c_{12} = .90 \qquad c_{78} = .99$$
$$h_1^2 = .81 \qquad h_8^2 = .96$$

Remembering that the complexities are simply the c_{ij}, we now have both our communalities and our complexities, and we can generate our theoretical correlations. The ρ_{ij} for the Two-Hand Coordination Test appear in Table A13, and, as the reader can see, we have a very good fit. Also, if the reader will check the parameters obtained by this method with those obtained by the communality method, he will see that they are essentially the same.

TABLE A13

ATTENUATED THEORETICAL CORRELATIONS FOR THE TWO-HAND COORDINATION TEST IN THE COVARIANCE METHOD

Trial	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1		.79	.75	.73	.72	.70	.69	.69
X_2			.88	.86	.86	.83	.82	.82
X_3				.90	.89	.86	.86	.86
X_4					.91	.88	.87	.87
X_5						.91	.90	.90
X_6							.92	.92
X_7								.94
X_8								

In conclusion, let me say a word or two about the comparative merits of the communality and the covariance methods. Ordinarily, the communality method is quicker and involves many fewer computations. But it can go awry. If the communalities of the variables happen not to coincide, even approximately, with the tentative values which are given us by one or another of our three methods of approach, we can waste an awful lot of time looking around for a fit. The covariance method, in comparison, is sure. If a fit exists. you can always get it with the covariance method.

Psychological Review
1962, Vol. 69, No. 4, 295–305

NORMATIVE AND IPSATIVE MEASUREMENT IN PSYCHOLOGY[1]

DONALD M. BROVERMAN

Worcester State Hospital, Massachusetts

In an earlier paper (Broverman, 1961) certain differences between factors obtained from standard *R* factor analysis methods (correlating tests) and *Q* factor analysis techniques (correlating people) were illustrated and attributed to the different types of scores used by the two techniques. The usual *R* analysis uses "normative" scores, or scores which reflect deviations of individual performances from the *group's* mean level of performance; while Q factor analysis typically uses "ipsative" scores, or scores which reflect deviations of particular performances from each *individual's* mean level of performance. Normative scores, then, are distributed about a norm characterizing a group, while ipsative scores are based on norms of individuals.

The present paper will attempt to show that normative and ipsative possibilities of measurement occur not only in factor analysis, but, also, in most efforts to quantify behavior, whether in purely theoretical terms or in actual experimental situations. Accordingly, the assumptions involved in each mode of measurement, and examples of these assumptions in current psychological theories will be pointed out. The paper will then examine the manner in which standard statistical procedures cope, or fail to cope, with these different aspects of behavior. With respect to this last point, the classical correlational procedures will be shown to confound normative and ipsative aspects of behavior. Hence, the final portion of this paper will present correlational techniques for dealing with these distinctions.

[1] Preparation of this paper was facilitated by the Dementia Praecox Research Project, Worcester State Hospital, and research grant (M-896), from the National Institute of Mental Health, United States Public Health Service. The writer prepared this paper while a United States Public Health Postdoctoral Fellow in Clinical Psychology.

Assumptions of the Normative Measurement Orientiation

The normative measurement orientation assumes that behavioral variations between individuals may be ordered on parameters which exist within the species' universe of variation. Thus, normative measurement procedures distribute scores about the norm of a group, sample, or population. In normative measurement, then, the behaviors of individuals are ordered on dimensions which transcend the individual, i.e., the behavior of a given individual is but a point or mark on a normative scale. This approach assumes that certain individual differences in behavior are due to an entity or trait which all members of the population possess in varying amounts. For example, Spearman (1927) assumes that individual differences in intelligence reflect the different amounts of mental energy possessed by different individuals.

It should be noted that intraindividual variation in parts of normative traits are not considered in normative measurement. Accordingly, less than perfect correlations between several manifestations of the same normative

trait are accounted for by assuming (*a*) that the surface behaviors showing the trait are phenotypically complex manifestations of more than one underlying genotypical, normative trait, and/or (*b*) that part of the discrepancy is due to measurement error. Thus, normative dimensions are treated as unitary, internally undifferentiated variables whose variation occurs across individuals.

ASSUMPTIONS OF THE IPSATIVE MEASUREMENT ORIENTATION

The ipsative measurement orientation, on the other hand, assumes that each individual is, himself, a universe within which variations in behavior occur. Accordingly, the different behaviors of an individual are ordered on dimensions or scales within each individual, e.g., the habit strengths of different responses in an individual may be scaled and compared. Particular behaviors of the individual, then, are assigned marks or scores on ipsative scales.

The ipsative measurement approach frequently assumes that relationships between psychic entities in an individual determine his behavior. For instance, Spence (1954) argues that when incompatible response tendencies are simultaneously aroused in an individual, the response possessing the strongest habit strength will be the one the individual will make.

It should be noted that often the ipsative orientation is not concerned with the unique aspects of single individuals as in the idiographic method advocated by Allport (1937). On the contrary, ipsative relationships are frequently ascribed in theory to an entire population, i.e., the rules or laws between internal psychic events may be generalized to all humans. Thus, it is not contradictory to seek commonly occurring ipsative relationships. Indeed, this is the goal of many personality theories.

The ipsative measurement orientation, then, is concerned with quantitative relationships between psychic functions within individuals. It frequently assumes that such intrapsychic relationships play an important role in the determination of behavior.

IPSATIVE MEASUREMENT PROCEDURES

A property of all scales, both ipsative and normative, is that the scale scores must be in the same units. Hence, in ipsative measurement, the various behaviors of a given individual must be expressed in terms of some common unit if scale comparisons are to be made. Numerous examples in psychological theory illustrating this procedure are available. For instance, Spearman (1927), who maintains that intelligence has certain subfunctions such as eduction of relationships, reproduction, etc., posits that a large expenditure of mental energy in one subfunction leaves less energy available for other subfunctions. In this example, the common unit of the different subfunctions is mental energy. The energy levels of the two subfunctions, therefore, may be scaled and compared. Similarly Freud (1935) posits conflict between different psychic institutions, e.g., id versus ego, the outcome of which depends upon their respective strengths. Once more the common denominator is some sort of mental energy.

Thus, in ipsative scaling procedures, different behaviors are reduced to, or expressed in terms of some common unit, and vary around a norm established within the individual. The norm, of course, is also expressed in these same units.

Different ways of establishing ipsative

norms within individuals exist. One method is by working with single, isolated individuals. For example, in Stephenson's (1953) *Q* sort each individual produces his own distribution of responses to a set of stimuli. The typical unit of variation in such *Q* sorts is the degree of "likeness to me." Another example would be measures of physiological reactivity under different conditions which are compared to a base-rate of that activity. In these examples, the norms are derived solely from a given individual's behaviors. What other individuals do has no effect upon the scales or norms so established. Hence, this may be called a "pure" ipsative measurement method.

A second way of establishing ipsative norms within individuals starts with normative considerations. For instance, the performances of a number of subjects might be normatively ranked for each test in a test battery. A given subject would then have a distribution of normative ranks—one rank for each of the tests in the battery. The subject's general level of performance for the battery, i.e., the subject's average normative rank standing, would then constitute the subject's ipsative norm and deviations of the subject's normative rank standings on particular tests from this ipsative norm would represent ipsative scores. The ipsative scores, in this case, are in units of deviations of group standings on various tasks from the subject's average group standing. Ipsative scores derived in this manner may be termed "double-centered" scores, i.e., they are centered or distributed first about a group norm, then about a norm of the individual.

Thus, unlike either the pure ipsative method or the normative approach, the double-centered ipsative approach takes into consideration both normative and ipsative aspects of variation.

MODES OF MEASUREMENT VERSUS PARAMETERS OF BEHAVIOR

Measurement operations refer to certain ways of assigning values to events or data on hypothetical parameters. Measurement-wise, however, one may develop several different techniques for establishing such values, or scores, on a given hypothetical parameter. Thus, one may distinguish between the concept of "modes of measurement" and the hypothetical parameters upon which these measurement techniques are intended to produce scores. The lack of equivalence between particular measurement or scaling procedures and the hypothetical parameters upon which score values are desired may be illustrated by considering a situation where certain modes of measurement are inadequate for the task at hand. For instance, we have previously mentioned that the normative measurement orientation does not consider intra-individual variations in subdomains of normative traits. Similiarly, the pure ipsative approach does not attempt to work with normative variations between individuals. However, if one accepts the notion that a given set of performance scores may reflect both normative and ipsative sources of variation, then both of the above measurement orientations become inadequate as procedures for assigning appropriate normative and ipsative scale values to the data. Rather, some method which takes into account both sources of variation should be employed. Thus, the hypothetical parameters on which the data are to be given scale values are seen to be conceptually independent or superordinate to the particular manner in which the various marks on the parameters are developed. Hypothetical normative and ipsative behavioral parameters, then, represent two distinct and sepa-

rate areas of concern for theories of behavior. It is not surprising that theories of psychology have progressed down both of these avenues of inquiry.

NORMATIVE AND IPSATIVE PSYCHOLOGICAL THEORIES

Modern psychology includes both normatively and ipsatively oriented theories. For instance, Krech's (1949) theory that differences in intelligence are due to different rates of cortical conductivity seems, essentially, a normative theory. The dimensions involved in Sheldon's (1944) body types also seem to be true normative parameters, as do the various traits of personality with which Allport (1937) has been concerned.

However, most theories concerning the mechanisms of psychological functioning assume that behavior is a function of relationships between psychological events occurring within a given individual. For instance, Helson's (1948) Adaptation Level Theory suggests that a perceptual response is determined by the relationship of the present stimulus to previously experienced stimuli. Similarly, Werner and Wapner's (1952) Sensory-Tonic Theory stresses the role of the relationship of sensory to tonic factors in determining perception. Thus, the above perceptual theories, many learning theories (Hull, 1943; Thorndike, 1932), and general personality theories (Freud, 1935; Lewin, 1935) all emphasize the importance of relationships between psychic events in determining behavior. These theories, then, are essentially ipsative in nature. Of course, many theories may include both normative and ipsative considerations. For instance, while Freud (1935) is primarily concerned with the relationships between psychic institutions, he also postulates normative, between individual differences in endowments of libidinal energy. Conversely, while Spearman (1927) is perhaps best known for the normative aspects of his theory, i.e., differences in intelligence are due to different amounts of mental energy, his theory also includes formulations about ipsative relationships between subfunctions of intelligence. These theories, and quite likely others, then, contain both normative and ipsative formulations.

It should also be noted that the normative-ipsative distinction is not represented in the language of individual differences. Individual differences are usually thought of only in a normative sense, i.e., in Anastasi and Foley (1949) differences between individuals are thought of as "a matter of degree. It is in this sense that individual differences are said to be quantitative rather than qualitative" (p. 59). However, this orientation tends to obscure the possibility that certain between-individual differences may be a function of varying ipsative, within-individual relationships, e.g., a psychoanalyst might argue that the differences between psychotics and normals are due to different intrapsychic relationships between ego and id. On the other hand, individual differences stemming from differing ipsative relationships can also be placed on normative scales, e.g., the above analyst might also say that one patient is more or less psychotic than another. Terminology is needed, then, to distinguish those individual differences which are due solely to variations on normative parameters from individual differences stemming from differing ipsative relationships. Hence, it is proposed that the purely normative individual differences be termed "first order" normative differences, while those individual differences stemming from ipsative, within-individual relationships might

be termed "second order" normative differences.

It seems, then, that there are two legitimate, but independent realms of inquiry for a general psychology of individual differences. These are (*a*) the establishment of those normative parameters of differences within the species that are not merely phenotypical expressions of ipsative internal relationships, and (*b*) the establishment of ipsative parameters of internal organization. As already demonstrated empirically, the generalizations that hold at one level of analysis can differ from those that hold at the other level (Broverman, 1960, 1961; Mandler, 1959).

NORMATIVE AND IPSATIVE MEASUREMENT IN STATISTICS

Relatively little explicit attention is currently given to the distinction between normative and ipsative measurement.[2] Consequently, numerous inconsistencies in the applications of these orientations exist. For instance, most psychologists seem to prefer normative approaches to research problems even though many, if not most, of present day psychological theories lean heavily on ipsative conceptualizations. The development of statistical procedures for separating these different aspects of behavioral data has also been uneven. Some statistical techniques discriminate between normative and ipsative aspects of data while others do not. The following two hypothetical examples, involving typical research problems and statistical procedures, illustrate (*a*) an appropriate separation of normative from ipsative aspects of data, and (*b*) a confounding of these two types of information.

[2] An outstanding exception is Cattell's (1944) paper concerning the use of normative and ipsative scores in psychology.

The first example involves assessing the significance of the difference between two means. An investigator wishes to determine whether motivating instructions improve performances on a particular task. Hence, the experimenter administers the task under a motivated and a nonmotivated condition to the same group of subjects. Assuming sequence effects are controlled, the experimenter might proceed to assess the significance of the difference between condition means via an analysis of variance technique. Since the experimental design employs repeated measurements, the total variance would first be partitioned into "between" and "within" individual sources of variation. The within individual variance would then be further divided into that associated with variation in conditions, and that which is not. After division by the appropriate degrees of freedom, the ratio of the two portions of within-individual variance would constitute an F test. If this test attains the required level of significance the experimenter is able to say that "The mean performance under the motivated condition is significantly different from the mean performance under the nonmotivated condition."

This statistical procedure clearly separates the normative from the ipsative aspects of the data. The between-individual variance reflects normative interindividual differences in task ability, independent of conditions, and is appropriately discarded since this variance is not the point of inquiry. The experimenter is not concerned with ability differences between subjects. He wishes only to know the effects of conditions on the subjects after such ability differences are controlled. The within-individual variance, which reflects ipsative or intra-individual variation, on the other hand, reveals that the experimental conditions

had significant effects upon performance when these effects were measured against each subject's own norm. This statistical procedure, then, sensitively discriminates between normative and ipsative aspects of data.

In contrast, correlational analyses typically give little attention to the normative-ipsative distinctions in behavior. Consequently, correlational procedures tend to confound inter-individual with intraindividual variations in performance scores as illustrated in the following example:

The experimenter wishes to know the correlations between several tasks. Accordingly, he administers each task to each subject in his subject sample and correlates the various scores.

Table 1 presents hypothetical standardized scores for 12 subjects on four tasks. Table 2 shows that the Pearson product-moment intercorrelations of these tasks are positive, ranging from .69 to .93. This standard correlational procedure accepts these correlations as the relationships between test performances, which, indeed, they are. However, in the light of the previous discussion, we know recognize that these relationships between test performances are a function of two sources of variation: (*a*) normative differences between individuals in general ability levels, and (*b*) within-individual ipsative variations in abilities on different tasks. The correlation coefficients in Table 2, then, are confounded resultants of these two sources of variation. This is the same confounding that the experimenter took pains to avoid when assessing the difference between conditions means. The analogue of the mean difference technique in correlational procedures would be to eliminate normative differences in general ability level between individuals before examining the intraindividual variations that occur as a function of differences in task requirements. The more precise experimental question would then be "What are the relationships between abilities on different tasks when differences in general ability levels between subjects are held constant or statistically controlled?" As will be shown below, these relationships may be quite dissimilar to those obtained from the usual confounded method.

TABLE 1

HYPOTHETICAL PERFORMANCE DATA STANDARDIZED SCORES

Subject	Test 1	Test 2	Test 3	Test 4	Mean standard score
1	67	62	58	55	60.5
2	62	67	55	58	60.5
3	58	55	62	67	60.5
4	55	58	67	62	60.5
5	53	51	49	47	50.0
6	51	53	47	49	50.0
7	49	47	51	53	50.0
8	47	49	53	51	50.0
9	45	42	38	33	39.5
10	42	45	33	38	39.5
11	38	33	42	45	39.5
12	33	38	45	42	39.5

TABLE 2

INTERCORRELATION OF TESTS

Test	Test 2	Test 3	Test 4
Test 1	93	69	69
Test 2		69	69
Test 3			93
Test 4			

Note.—Decimals have been omitted.

Our problem is to separate the variance attributable to normative differences between subjects from that due to within-individual variations in ability to perform different tasks.

Normative variance between subjects occurs when some subjects tend to be generally above the group mean, while others tend to be below this reference norm. However, if the task per-

TABLE 3

HYPOTHETICAL PERFORMANCE DATA TRANSFORMED INTO IPSATIVE SCORES
$(X_i - \bar{X}_i)$

Subject	Test 1	Test 2	Test 3	Test 4
1	6.5	1.5	−2.5	−5.5
2	1.5	6.5	−5.5	−2.5
3	−2.5	−5.5	1.5	6.5
4	−5.5	−2.5	6.5	1.5
5	3.0	1.0	−1.0	−3.0
6	1.0	3.0	−3.0	−1.0
7	−1.0	−3.0	1.0	3.0
8	−3.0	−1.0	3.0	1.0
9	5.5	2.5	−1.5	−6.5
10	2.5	5.5	−6.5	−1.5
11	−1.5	−6.5	2.5	5.5
12	−6.5	−1.5	5.5	2.5

formances of each subject are expressed as ipsative scores, or deviations from each subject's mean level of performance, then all subjects will have the same status with respect to the new reference norms, i.e., the sum of the deviations above each ipsative norm will equal the sum of the deviations below for each subject. Hence, normative differences between subjects are, in effect, removed via ipsatization of performance scores. Correlation of the ipsative scores will then reveal the intrapsychic relationships between abilities on different tasks, free from the influence of differences in general ability between subjects.

The mean standard score of each subject is shown in Table 1. Table 3 shows the ipsative scores $(X_i - \bar{X}_i)$ of each subject on each task. Table 4 shows the intercorrelations of these ipsative scores. It will be noticed that some of the correlations are now negative, others positive. Thus, controlling normative interindividual differences in general ability levels considerably alters the correlational relationships between tasks, and, of course, any conclusions that might be drawn from these data. Instances where such controls have actually affected the results of studies have been reported by Mandler (1959) and Broverman (1960, 1961).

It seems clear then, that correlational analysis, a classical research tool in systematic psychology, has suffered from a lack of sensitivity to the normative-ipsative distinctions in behavior. Standard R factor analysis technique, which is an elaboration of the correlational method, suffers in the same manner, i.e., the factors extracted by these techniques are confounded expressions of normative and ipsative variation. The confounding in R factor analysis can be overcome in the same manner as illustrated above, i.e., by ipsatizing the various performance scores before starting correlational and factoring procedures. As in the above correlational example, factors derived from ipsative scores will frequently have large negative loadings or correlations with tests. A discussion of the numerical mechanics underlying this phenomenon may be found in Broverman (1961). That paper also pointed out that an ability factor defined by both positive and negative loadings can be viewed as a "choice point" in development since at some point in each individual's development a decision must be made as to whether he is to specialize in one or the other of the two classes of behaviors defining the factor. Other investigators (Cattell & Baggaley, 1956; Sweney, 1959) have interpreted this specialization or directional

TABLE 4

INTERCORRELATIONS OF IPSATIVE SCORES

Test	Test 2	Test 3	Test 4
Test 1	58	−79	−79
Test 2		−79	−79
Test 3			58

Note.—Decimals have been omitted.

aspect of ipsative scores as an index of motive strength.

RELATIONSHIP BETWEEN NORMATIVE PARAMETERS AND PARAMETERS DERIVED FROM DOUBLE-CENTERED IPSATIVE SCORES

The relationship between dimensions derived from ipsative scores distributed about a normative base-line, and the normative dimension itself is worthy of particular note. Such ipsative dimensions are internal subdivisions or differentiations of the larger normative dimension. Again, Spearman's (1927) theory of intelligence illustrates this point. Spearman argues that intelligence varies in units of mental energy. Normatively, then, the subjects differ from each other in terms of the amount of mental energy possessed by each. However, within intelligence, as mentioned earlier, Spearman posits subfunctions of eduction, reproduction, etc., the strength of which are also measured in terms of mental energy and which vary ipsatively within each individual. These subfunctions, then, are internal differentiations within the larger normative trait of intelligence.

This same part-whole relationship between ipsative and normative factors is obtained when the ipsative dimensions are empirically derived from double-centered ipsative scores via factor analysis. For instance, a subject's average rank performance for a battery of cognitive tests might be interpreted, normatively, as an indication of his general level of intelligence. On the other hand, the factors emerging from an analysis of the ipsative scores distributed about the average rank standings of subjects would represent internal differentiations of that normative trait.

This part-whole relationship of ipsative to normative factors may be contrasted with the relationships between factors produced by the usual normative *R* factor analysis procedures. Factors produced by this latter method are discrete, and juxtaposed externally to each other. Even though normative *R* factors may be correlated with each other, one normative factor is not thought of as a subdivision of another, as in the double-centered ipsative approach.

As pointed out by Stephenson (1953), factor analysis is typically used in an exploratory way to determine the various interdependencies between psychological functions or behaviors. When statistical procedures are used in this exploratory manner, one tends to allow the model embodied in the statistical procedures to guide one's conception of the organization of the behavior reflected in the data. Thus, the 30 years of research employing normative *R* factor analysis, with its model of externally juxtaposed factors, has left its mark on present conceptions of mental functioning, e.g., Burt's *The Factors of the Mind* (1946), Thurstone's *Primary Mental Abilities* (1938). While the part-whole model embodied in double-centered ipsative measurement is not new on the psychological scene—witness Spearman's (1927) theory of intelligence—a clearer recognition of its properties may exert a correspondingly greater influence upon our conceptions of the organization of behavior.

DETERMINATION OF APPROPRIATE NORMATIVE BASE-LINES

A major problem in double-centered ipsative measurement is the identification of appropriate normative base-lines from which to develop ipsative scores. So far, we have assumed that a single normative base-line may be used for a given matrix of data. This assumption is implicit in the analysis of variance repeated measurements de-

sign. However, it is possible that more than one normative variable, hence, more than one base-line, is actually embedded in a given matrix of data. In this case, certain behaviors ought to be ipsatively measured against one base-line, others against another. Hence, a method for objectively establishing the presence of normative variables is needed. This problem may be resolved via the following considerations:

Any single performance score may be thought of as a resultant of normative and ipsative sources of variation. Consider, now, a situation where several tests all reflect a single normative trait, but some of these tests also reflect one ipsative dimension, other of these tests reflect a second ipsative dimension, still others reflect a third ipsative dimension, etc. Each test score of a given subject, then, consists of a constant representing his status on the normative dimension and a variable representing the standing of the task on a particular ipsative dimension. Since this normative constant varies between subjects, it tends to produce positive relationships between tests. The ipsative components of the various performance scores, on the other hand, will either correlate positively when they reflect the same ipsative dimension, or not correlate when they reflect different ipsative dimensions. Still other ipsative components may be inversely related and work towards negative relationships. Thus the normative components tending to produce positive relationships between tests will summate with those positive ipsative covariations, but will be weakened or canceled out by those independent and negatively varying ipsative components. The first order (simple structure) factors extracted from performance scores, then, will be those tests whose ipsative components have positive relationships augmented by their positively covarying normative components.

Since each of these factors have the same normative components, only the variation of the ipsative components can be acting to produce the different factors. Also, since each factor has the same normative component, these factors should correlate with each other. Accordingly, the factor intercorrelations ought to produce a second-order factor reflecting the presence of the normative trait. Of course, if the data actually contained more than one normative trait, then more than one second-order factor would be obtained. The number of second-order factors obtained from a factorial solution of performance data, then, indicates the number of normative dimensions present in the data. In turn, each normative dimension calls for its own base-line from which to assess ipsative variations of the tests included in the second-order factor. Thus, the mean group standing of all tests included in each second-order factor would be obtained for each subject. Deviations of individual tests from these means would then reflect ipsative variations within a given normative dimension. Other methods of obtaining appropriate base-lines for ipsative oriented research problems may be found in Broverman (1960).

CONCLUSION

It may be well to conclude by attempting to see the issues discussed in this paper in a broader perspective. It seems that the measurement of behavior is more complicated than previously suspected. One wonders why this should be so; why one can not proceed to a more straightforward measurement of behavior. The reasons seem at least three-fold:

First, human behavior is so bewilderingly varied and multifaceted that

few psychologists are willing to stay at the level of raw, uncategorized, observations of behavior. Such an enterprise would produce an infinity of parameters, variables, and scales. Rather, one typically seeks to reduce a large number of different behaviors to a smaller number of common underlying parameters. The pertinent variables of behavior, then, are not accessible for direct measurement. One must fractionate, peel off, and discard the surface aspects of different behaviors in order to uncover and obtain an estimate of the underlying variables. Thus, in measuring behavior, phenotypical surface appearances are discarded in favor of inferred genotypical variables.

A second aspect of human behavior which contributes to the complexity of its measurement is that any given behavior can simultaneously represent more than one underlying variable. A given behavioral act may be a composite expression of many different variables. Thus, one must separate not only phenotype from genotype, but also genotype from genotype.

Finally, we must recognize that all human behavior potentially participates in two different realms of human variation, i.e., species and individual. Each behavioral act is, potentially at least, a simultaneous expression of the human species' as well as of the individual's repertoire of variation. It is this dual aspect of behavior which produces the particular normative-ipsative model elaborated in this paper. Failure to represent this fact in measurement and statistical operations must lead to a confounding of these two different aspects of human existence and variation. If one accepts the proposition that the laws or regularities which characterize variations in behavior between individuals are not the same as those which characterize variations in behavior within individuals, then such confounding is to be studiously avoided. Hence, an orderly inquiry into the nature of human behavior requires a stringent separation of the between-individual species parameters from the ipsative intraindividual organizational aspects of behavior in order that the separate relational networks of each realm may be established. While this may involve considerable more work than previous methods, the results may be proportionately more rewarding.

REFERENCES

ANASTASI, ANNE, & FOLEY, J. P. *Differential psychology.* New York: Macmillan, 1949.

ALLPORT, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

BROVERMAN, D. M. Cognitive style and intraindividual variations in abilities. *J. Pers.*, 1960, **28**, 240–256.

BROVERMAN, D. M. Effects of score transformations in *Q* and *R* factor analysis techniques. *Psychol. Rev.*, 1961, **68**, 68–80.

BURT, C. L. *The factors of the mind.* London: London Univer. Press, 1946.

CATTELL, R. B. Psychological measurement: Ipsative, normative and interactive. *Psychol. Rev.*, 1944, **51**, 292–303.

CATTELL, R. B., & BAGGALEY, A. R. The objective measurement of attitude motivation: Development and evaluation of principles and devices. *J. Pers.*, 1956, **24**, 401–423.

FREUD, S. *A general introduction to psychoanalysis.* New York: Liveright, 1935.

HELSON, H. Adaptation-level as a basis for a quantitative theory of frames reference. *Psychol. Rev.*, 1948, **55**, 297–313.

HULL, C. L. *Principles of behavior: An introduction to behavior theory.* New York: Appleton-Century, 1943.

KRECH, D. Notes toward a psychological theory. *J. Pers.*, 1949, **18**, 66–87.

LEWIN, K. *A dynamic theory of personality: Selected papers.* New York: McGraw-Hill, 1935.

MANDLER, G. Stimulus variables and subject variables: A caution. *Psychol. Rev.*, 1959, **66**, 145–149.

SHELDON, W. H. Constitutional factors in personality. In J. McV. Hunt (Ed.),

Personality and the behavior disorders. Vol. 1. New York: Ronald, 1944.

SPEARMAN, C. *The nature of intelligence and the principles of cognition.* London: Macmillan, 1927.

SPENCE, K. W. Current interpretations of learning data and some recent developments in stimulus-response theory. In, *Kentucky Symposium: Learning theory, personality theory, and clinical research.* New York: Wiley, 1954. Ch. 1.

STEPHENSON, W. *The study of behavior.* Chicago: Chicago Univer. Press, 1953.

SWENEY, A. B. The objective measurement of motivation in children. *Amer. Psychologist,* 1959, **14**, 355. (Abstract)

THORNDIKE, E. L. *The fundamentals of learning.* New York: Teachers College, Columbia University, 1932.

THURSTONE, L. L. *Primary mental abilities.* Chicago: Chicago Univer. Press, 1938.

WERNER, H., & WAPNER, S. Toward a general theory of perception. *Psychol. Rev.,* 1952, **59**, 324–338.

(Received October 24, 1960)

Psychological Review
1962, Vol. 69, No. 4, 306-328

FRUSTRATIVE NONREWARD IN PARTIAL REINFORCEMENT AND DISCRIMINATION LEARNING: SOME RECENT HISTORY AND A THEORETICAL EXTENSION [1]

ABRAM AMSEL

University of Toronto

The overall objective of this paper is to examine the current status of a concept of frustrative nonreward in behavior theory. The examination takes two forms: a general discussion and survey of recent work on motivating and inhibiting properties of nonreward, followed by a theoretical extension of frustrative-nonreward theory to certain prediscrimination phenomena.

In the first part of the paper a brief historical account will be presented of recent changes toward an "active" conceptualization of nonreward, particularly in neo-Hullian theory, the current status of such an active nonreward concept in explanations of partial reinforcement effects will be set in a broader framework of partial reinforcement theories, and some recent data will be presented in support of the idea that discrimination learning involves frustrative-nonreward effects.

The second more theoretical portion of this paper stems from an earlier frustrative conceptualization of nonreward (Amsel, 1958). It represents an attempt to extend frustrative-nonreward theory to situations in which discrimination learning is preceded by various amounts and kinds of "prior experience." These prior experiences will be termed "prediscrimination treatments" and will involve training to approach in the presence of one or both (separately) of the eventual discriminada, under conditions of partial or of continuous reinforcement.

[1] A shorter version of this paper was presented to the Psychology Colloquium of the University of Illinois in May 1961. The preparation of this paper and much of the experimental research to which it refers were supported by grants (G-5527 and G-13895) from the National Science Foundation.

History

Hull's adoption of what he termed the Mowrer-Miller hypothesis as a basis for a two-factor theory of inhibition (Hull, 1943) was, at the same time, the acceptance of a "passive" conception of nonreward and rejection of a more active conception which he and others had held. Certainly, Pavlov's nonreinforcement, producing internal inhibition, was an active conception. Spence's (1936, 1937) theory of discrimination learning, clearly in the Pavlovian tradition, also assigned special inhibitory properties to nonreinforcement. In some of his more recent systematic writing, Spence (1956) points out that his position on the Mowrer-Miller hypothesis was never in agreement with Hull's.

Emotional-motivational effects of nonreward, as they relate to learning, were described in an article by Miller and Stevenson (1956), which attributed the agitated behavior of rats during extinction to effects of nonreinforcement carried over from one trial to the next. This is an example of an active motivational conception of nonreward. A similar emotional-motivational interpretation of nonreward can be found in Skinner (1938).

In Hull's 1943 theory, which domi-

nated the psychology of learning in the forties, nonreinforcement was an event without direct inhibitory or motivational effect as it was also in the early mathematical-model treatments of learning (Bush & Mosteller, 1951; Estes, 1950). By about 1950, however, expressions of dissatisfaction with the passive theoretical status of nonreward were emerging. These expressions were in theoretically oriented experiments and in theoretical papers. Most of the early defections from 1943 Hull, in this connection, were by persons close to the Hullian point of view. There are suggestions that Hull, himself, was getting ready to change his position on this shortly before his death (Hull, 1952, Ch. 5).

Rohrer's (1949) treatment of extinction in terms of a "frustration drive" was one of the first expressions of a desire to give nonreinforcement a more active role in inhibition.[2] However, while Rohrer's reference to frustration was specifically related to the operation of nonreinforcement, it provided no particular mechanism for frustrative inhibition. At about the same time, the well-known experiments of Virginia Sheffield (1949, 1950) appeared, relating partial reinforcement acquisition and extinction effects to the spacing of trials and containing suggestions of both frustration drive and frustrative inhibition. These articles and another by Stanley (1952), along with a late experimental paper by Hull and others (Hull, Livingston, Rouse, & Barker, 1951) and Hull's (1952) last book, all contained at least the germ of a change toward a more active conception of nonreinforcement. An experimental paper by Denny and Dunham (1951) offered inhibition based on nonreinforcement (frustration) as an alternative to reactive inhibition to account for differential nonreinforcement effects in a T maze.

Brown and Farber (1951) outlined a theory in which nonreward was one of several antecedents to frustration. In this paper, the major emphasis was on the definition of frustration in terms of competing response tendencies and the authors were concerned only with the motivational (drive) properties of frustration, and not with inhibitory effects. In the same year, a paper by Amsel (1951) made the point that fractional anticipatory frustration (r_F), which is the classically conditioned form of the frustrative reaction to nonreward, provides a mechanism for conceptualizing the active properties of nonreinforcement and should be regarded as a determiner of inhibitory effects. And, at about the same time, Seward (1951) published a theoretical paper which mentioned, without elaboration, the possible functional properties of an anticipatory frustration factor.

The remainder of the decade, 1950 to 1960, has seen a number of additional experimental and theoretical attacks on the problem of nonreward. On the one hand there are experiments and theoretical treatments by investigators in the Hullian tradition of behavior theory which have sought to extend and revise Hull's theory in regard to the status of nonreward. The other hand holds a number of published studies representing a variety of theoretical and nontheoretical interests which have contributed information about frustrative nonreward.

[2] It is questionable that "inhibition" is the appropriate and correct term to use in regard to response decrement related to frustrative nonreward. Better terms would be "interference" or "competition," since frustrative inhibition cannot be regarded as directly affecting the excitatory tendency out of which it grows. It is rather a new excitatory tendency which competes with the older one.

Much of this work has been reviewed earlier (Amsel, 1958; Lawson & Marx, 1958); but some of it is very recent. The experiments fall into three categories: (*a*) demonstrations of the motivating effects of frustrative nonreward, i.e., the carried-over, enhancing effect of nonreward on the vigor of immediately following behavior. This has been termed the Frustration Effect (FE). Several studies, by far the greatest number, have been directed not only at demonstrating the effect but also at examining the variables determining its occurrence and strength (Amsel, Ernhart, & Galbrecht, 1961; Amsel & Hancock, 1957; Amsel & Penick, 1962; Amsel & Roussel, 1952; Bower & Stocks, 1960; Longstreth, 1960; Marzocco, 1951; Penney, 1960; Roussel, 1952; Seward, Pereboom, Butler, & Jones, 1957; Wagner, 1959). (*b*) The second category consists of experiments which are mainly concerned with the frustration drive stimulus, but which also touch on frustration reduction as reinforcement (Amsel & Prouty, 1959; Amsel & Ward, 1954; Tyler, Marx, & Collier, 1959). (*c*) And finally, there are a few studies which deal with the inhibitory properties of frustrative nonreward. The inhibitory mechanism is fractional anticipatory frustration (r_F), a classically conditioned form of the goal response to frustrative nonreward. The mechanism and its theoretical properties have been identified in detail (Amsel, 1958; Spence, 1960), and experiments testing implications of such a mechanism are available (Bower, 1961; Goodrich, 1959; Haggard, 1959; Wagner, 1961a, 1961b). Of particular importance in this connection are a series of experiments by Wagner (1961b) which go far toward establishing the proposition that frustrative nonreward has many of the properties of punishment, and that r_F operates in many respects like fear. These experiments would support such statements as (*a*) the greater the strength of frustration the faster the extinction (avoidance) of a nonrewarded response; (*b*) cues paired with frustrative nonreward acquire motivational properties; and (*c*) the cessation of cues previously paired with frustrative nonreward serves to reinforce a new response.

There are several other experiments on nonreward (or reward prevention) as a variable in learning which provide support for nonreward as a factor in inhibition. For example, studies by Lambert and Solomon (1952) and by Adelman and Rosenbaum (1954) related extinction to the blocking of instrumental behavior at various distances from the goal. An experiment by Adelman and Maatsch (1955) found that resistance to extinction depends to some extent on what the subject is allowed to do when it finds no reward in the goal box, i.e., is frustrated. The suggestion is that resistance to extinction is high when the subject makes a response to frustration which removes it from the frustrating situation rather than one which keeps it in the apparatus. Ferster (1957, 1958) found that stimuli signaling "time out" from positive reinforcement acquire aversive properties. He also showed that the withdrawal of a positive conditioned reinforcer had the functional properties of punishment, suppressing the rate of responding much like the presence of a negative reinforcer. These experiments point clearly to the inhibitory (aversive) properties of nonreward.

Some recent experiments, which show decreasing resistance to extinction with increasing numbers and magnitudes of reinforcement, must be regarded as supporting a conception of nonreward as an active factor. For

example, Hulse (1958) and Armus (1959) found that larger magnitudes of reinforcement in acquisition are followed by faster extinction. Very recent reports by Reynolds and Siegel (1961), by Pavlik (1961), and by Wagner (1961a) were to the same effect. North and Stimmel (1960) have shown that rats reinforced 90 or 135 times extinguish faster than those reinforced 45 times. Similar results have been reported by Reynolds, Richter, and Carlock (1960). They found faster extinction following .5- than following .1-gram reward, and faster extinction after 66 acquisition trials than after 6, 12, or 30 trials. This kind of result is quite in accord with results from studies of overlearning and discrimination reversal, which we will look at later. Together, they are as incompatible with early Hullian theory, or any other theory requiring monotonicity and positive relationships in these variables (e.g., the various mathematical approaches), as were the early partial reinforcement findings. We will come back to some of these experiments later, and my discussion of partial reinforcement effects will show why a theory which includes a frustrative nonreward hypothesis must predict such results.

There are by now several experiments which can be taken to indicate that nonreinforcement is more important than reinforcement in discrimination. Chronologically, these are studies by Fitzwater (1952), Grove and Eninger (1952), Eninger (1953), Shoemaker (1953), Cantor and Spiker (1954), Grice and Goldman (1955), Birch (1955), and Solomon (1956), to mention some of the earlier ones. A very recent study which makes this same point has been reported by Lachman (1961). This is not the place for a detailed description of these studies; however, the burden of evidence is that avoidance of nonreward is a more powerful factor in discrimination than approach to reward; and this is, of course, very relevant to a point of view which makes nonreward a determiner of inhibition. Perhaps Harlow and Hicks (1957) were reacting not only to their own experimental findings but also to the weight of this kind of evidence when they argued for a "uniprocess" rather than a "duoprocess" theory of discrimination learning, the single process being inhibition. The weakness of their argument is that the negative process (r_F, in my terms) may well depend on the prior existence of a positive process (see Amsel, 1958; Spence, 1936, 1937), and one is back to two processes again.

Other examples of a return to an active conception of nonreward outside of the Hull-Spence camp can be found in the recent work of Mowrer and Estes. Mowrer, in his *Learning Theory and Behavior,* makes frequent use of the notion of nongoal events arousing "frustration" or "anger," in addition to "disappointment" and expresses agreement with the position I favor in regard to frustrative goal events (Mowrer, 1960, p. 409). Estes (1959) has recently acknowledged that statistical learning theory must cope with the problem of an active negative process based on nonreward in these words:

> . . . I have found that in the naive animal an unreinforced trial produces no apparent change in response probabilities, but that after a series of reinforced runs to a given side of a T maze, the introduction of nonreinforcement yields a decrement in probability of response to that side. . . . Apparently under partial reinforcement schedules, the effect of nonreinforcement varies between these extremes . . ." (p. 417).

From this brief sketch of the recent history of the concept of nonreward,

we move to a description of the role of active conceptions of nonreward in interpretations of noncontinuous (partial) reinforcement effects.

NONREWARD IN INTERPRETATIONS OF PARTIAL REINFORCEMENT

In his 1958 chapter on learning, Lawrence differentiates between "intertrial" and "intratrial" interpretations of the Partial Reinforcement Effect (PRE)—the finding that partial or intermittent reinforcement in learning leads to greater resistance to extinction than does continuous reinforcement. Intertrial explanations attribute the PRE in one way or another to immediately carried-over traces of stimulation from one trial to the next in acquisition. According to such interpretations, it is on the basis of such traces that acquisition is more clearly discriminated from extinction by continuously than by partially reinforced subjects. Intratrial explanations ascribe the PRE to some kind of learned mechanism developing on each trial in acquisition and having its effect, associatively, later in extinction. There have been two classes of intertrial explanations. In both cases the classes can be termed Cognitive-Expectancy (C-E) interpretations, and Stimulus-Response (S-R) interpretations.

Generally speaking, the intertrial interpretations were an earlier development than the intratrial interpretations. Early C-E interpretations were the "common sense" expectancy hypothesis which emerged out of the early important experiments of Humphreys (e.g., 1939, 1940) and the discrimination hypothesis developed and tested by Bitterman and his associates (e.g., Bitterman, Fedderson, & Tyler, 1953). On the S-R side, the major intertrial explanation (apart from the "response unit" interpretation which applies to chained or highly massed responses but not to discrete trial learning) was the Hull-Sheffield hypothesis tested by Sheffield (1949). Weinstock's (1954) demonstration that partial reinforcement effects were not eliminated even when trials were separated by 24 hours, and an experiment by Wilson, Weiss, and Amsel (1955) disconfirming the earlier results of Sheffield helped to set the stage for, and indeed to provide, intratrial explanations of, the PRE. Much later, in fact very recently, there has been strong contraindication for a discrimination hypothesis in the work of Jenkins (1961b) and of Theios 1962). Jenkins, using pigeons in a Skinner-type situation, and Theios, using rats in a runway, have found that the PRE is sustained when a large number of continuous reinforcements are interpolated between discrete trial partial reinforcement and extinction. This means that a large number of continuous reinforcements immediately before extinction and following partial reinforcement does not neutralize the partial reinforcement effect. Since the basis for discriminating acquisition from extinction is now the same for partially as for continuously reinforced subjects, and the PRE persists, an interpretation is required, as Theios indicates, in terms of "relatively permanent effects of nonreinforcement which can be sustained through blocks of continuous reinforcement."

Intratrial explanations of the PRE concern themselves in one way or another with processes following nonreward. Until recently, the intratrial explanations were the exclusive property of S-R psychologists. Weinstock's (1954) contiguity-interference hypothesis was perhaps the earliest of these, but was followed in short order by a variety of neo-Hullian hypotheses. All of these are what I would term conditioning-expectancy hypothe-

ses as opposed to cognitive-expectancy hypotheses of the PRE. All depend upon constructs of the anticipatory goal response type, representing responses which develop in strength within trials. There are three neo-Hullian interpretations of the PRE, all of which were first proposed to account for particular experimental findings. The first emerged out of a study by Wilson, Weiss, and Amsel (1955) and was subsequently elaborated and tested in some detail (Amsel, 1958). It attributes the PRE to the evocation of frustration by nonreward in partial reinforcement, to the conditioning of anticipatory frustration ($r_F - s_F$) to cues in the instrumental sequence, and to the association of s_F with approach responses in acquisition. Logan's interpretation of the PRE (Logan, Beier, & Kincaid, 1956) is that it depends on the extinction of r_G to "postreinforcement-time cues." Kendler and his associates (Kendler, Pliskoff, D'Amato, & Katz, 1957) have offered an interpretation similar to the one I have favored, except they choose a neutral designation of nonreward effects in acquisition and refer to anticipatory nonreward and not to anticipatory frustration. Spence (1960) has recently come out in favor of the same kind of interpretation and has employed the r_F construct to account for acquisition phenomena in partial reinforcement. Consequently, it seems safe to say that a prominent interpretation of the PRE in neo-Hullian writings is a conditioning-expectancy position with identifies fractional anticipatory frustration as well as fractional anticipatory reward.

Recently, Festinger (1961) has called attention to the importance of "insufficient rewards" in a cognitive-expectancy (cognitive dissonance) interpretation of the PRE. Cognitive dissonance (Festinger, 1957) is a hypothesis which recognizes the active properties of nonreward. In its use of delay of reward procedures to enhance attractiveness of a goal in a delay of reward situation, Festinger's treatment of insufficient rewards is very similar in conceptualization to some earlier work of Olds (1953, 1956). On the basis of experiments with children, Olds concluded that the reward value of a secondary rewarding stimulus (S^r) could be increased—and hence the strength of approach responses leading to S^r could be increased—by delaying the presentation of S^r *after* overlearning. He showed that tokens which could ultimately be exchanged for more primary reward increased in reward value when their presentation was delayed, if these delay trials followed a number of trials in which their presentation was immediate. Olds suggested that such a procedure constitutes "practice at wanting" and that such practice at wanting increases the value of that which is wanted (S^r). Festinger would seem to be taking the same position in his animal research. In one of several experiments dealing with dissonance in rats, he employs a double runway in which a start box is separated from a mid box by an alley, and the mid box is separated from an end box by another alley. Subjects are delayed in the mid box in the manner of Holder, Marx, Holder, and Collier (1957) and Wagner (1959), and the finding is that the subjects delayed in the mid box, before being allowed to run to the end box and find food, will continue to run to the mid box longer on test trials than the subjects always fed in the mid box as well as in the end box. Festinger's (1961) interpretation is that the delayed subjects develop dissonance as a result of the delay, and that the "dissonance tends to be reduced by developing some extra preference about something in

the situation. The existence of this extra preference leads to the stronger inclination to continue running during extinction trials" (p. 9).

The position of Festinger, like the earlier one of Olds, is clearly that nonreward (or delay of reward) is an active factor determining an increased want or attraction. However, there seems to be an important difference in what increases in attractiveness between Festinger's earlier treatments of dissonance in humans (Festinger, 1957) and his application of the concept to the explanation of the PRE in rats (Festinger, 1961). While specificity is not always a characteristic of cognitive theory, it does seem as though the 1957 treatment (and even parts of the 1961 paper) did specify that the violation of a cognitive expectation produces dissonance which is resolved by upgrading the value of the previous goal object. However, in the treatment of "insufficient rewards" in terms of the prototype experiment with start box, mid box, and end box, dissonance is said to develop in the mid box, where the subject never has found food; and dissonance is reduced "by developing a preference for something about the *place* where the dissonance was introduced . . . by developing some liking for the place where they were not rewarded . . ." (p. 10). This latter position is used to explain, in terms of dissonance, why the subjects never fed in the mid box run longer to the mid box on nonreward test trials than the subjects always fed there before. The same thinking is then carried over to explanation of the PRE. In terms of the S-R analysis, the subjects have been trained to approach the mid box of Festinger's situation despite anticipatory frustration (s_F) cues because, in the period prior to testing, they were always finally fed in the end box. In C-E terms the subjects developed an extra preference for the nonreward situation —the mid box in which they had never been rewarded. This would leave to be answered such questions as: Why does successive nonreward result in extinction? And, how do the subjects learn to discriminate cues leading to reward from cues leading to nonreward?

The S-R treatment of nonreward effects in partial reinforcement (Amsel, 1958; Kendler, Pliskoff, D'Amato, & Katz, 1957; Spence, 1960; Wilson, Weiss, & Amsel, 1955) does not suggest that nonreward (or delayed reward) increases "wanting" in Olds' terms, or "preference about something in the situation" in Festinger's. This type of neo-Hullian approach is a description of a mechanism operating through anticipatory nonreward which can lead to either rapid extinction of behavior or to increased resistance to extinction, and which can either facilitate discrimination learning or retard it. I will claim that neither increased "wanting" nor "preference"—in fact, no "common sense" cognitive interpretation—can account for all of the effects which nonreward seems to have.

The difference between the two positions, in the cognitive language of Olds and Festinger, is as follows: While the C-E position has nonreward *increasing* attractiveness of the nonreward situation, the S-R position implies a *decrease* in attractiveness. The partial reinforcement effect is, according to Festinger, due to increased attractiveness. According to my position it is the result of training to persist in behavior despite decreased attractiveness. In some other respects the positions are similar: both make nonreward the important determinant

of the effect, and both stress the importance of developing (intratrial) associative tendencies during acquisition as determinants of ultimate resistance to extinction. Neither necessitates an assumption that the subject discriminates acquisition from extinction (Bitterman, Fedderson, & Tyler, 1953).

The details of the S-R position which I favor have been presented elsewhere (Amsel, 1958) in the form of a sequence of hypotheses identified with four stages of practice in discrete trial partial reinforcement or in discrimination learning (the first three stages are essentially the same for both). A fifth hypothesis which applies to discrimination learning has been added, and this will be important for some data which follow. The essence of the position is outlined below in relation to an apparatus we have been using. It was designed to differentiate the immediate motivating (energizing) effects of frustration exemplified by the so-called Frustration Effect (FE) from fractional anticipatory frustration (r_F), which represents the inhibiting feature of frustration. The apparatus is essentially, two runways in series, either in a straight line (e.g., Amsel & Roussel, 1952) or in the form of an L (e.g., Amsel & Hancock, 1957). The apparatus has a start box (SB), a first runway (Runway 1), a first goal box (G1), a second runway (Runway 2), and a second goal box (G2) in successive arrangement. Reward and nonreward are manipulated in G1, and the apparatus permits a separation of the frustrated response and changes in running speed indicating r_F (measured in Runway 1), from the frustration motivated response and changes indicating the FE (measured in Runway 2).

The sequence of hypotheses in terms of five stages of practice is as follows:

1. In Stage 1, $r_R - s_R$ [3] is developing in Runway 1 with early rewards, and nonreward (in G1) has no particular effect, as measured by the FE in Runway 2.

2. After the development of $r_R - s_R$ for the Runway 1 response, nonrewards elicit frustration, measured by the FE in Runway 2. This is Stage 2.

3. When nonrewards elicit frustration, the cues in Runway 1 previously evoking r_R now also begin to evoke r_F. In Stage 3 these antedating goal response tendencies are temporarily in competition.

4. Since r_R and r_F cannot be elicited separately by differential cues in partial reinforcement, as they can in the latter stages of discrimination learning, and since the temporary conflict in partial reward training is resolved in favor of running to the intermittently rewarding goal box, s_F becomes associated with the instrumental approach response in Stage 4 of partial reinforcement training, providing the mechanism for the partial reinforcement effect. When extinction is carried out, partially reinforced subjects have been trained to respond (approach in the presence of antedating,

[3] I employ the term, $r_R - s_R$, in the same manner as it is used in an earlier development, and for the same reason: ". . . $r_G - s_G$ is a general term covering all types of antedating conditioned response. Separately, these might be designated $r_R - s_R$, fractional anticipatory reward; $r_P - s_P$, fractional anticipatory punishment, usually termed fear or anxiety; and $r_F - s_F$, fractional anticipatory frustration" (Amsel, 1958, p. 102). The meaning of $r_R - s_R$ is, however, not different from $r_G - s_G$ as Spence (1960) uses it in a similar connection. All that is intended here is a clarification of terminology in light of the fact that goal events may be positive (such as to elicit approach) or negative (such as to elicit avoidance).

frustration produced stimuli), whereas consistently reinforced subjects have not. In discrimination situations there is a basis for the differential elicitation of approach and avoidance by s_R and s_F respectively in Stage 4.

5. Following discrimination, when r_F consistently precedes nonreward, the FE following nonreward should grow weaker, since nonreward is defined as frustrating only in the presence of r_R.

I have been suggesting that in a very important area of the analysis of behavior, the explanation of persistence of behavior in the face of nonrewards attributable to partial reinforcement, the role of nonreward as an active process has emerged in both cognitive-expectancy and in S-R conditioning interpretations. It would certainly seem to me that this is an area where the clearer, better worked out S-R language and conceptualization might provide a handle for a more workable expectancy theory.

There is, in addition, the matter of a basic difference between the cognitive and S-R interpretations of the PRE: the former says nonreward increases attractiveness; the latter says nonreward decreases attractiveness. Both positions can account for the PRE; both can account for faster extinction following large rewards than following smaller rewards; both can account for faster extinction following many rewards than following fewer rewards. However, a cognitive theory such as Festinger proposes would have difficulty with a finding of Hulse (1958) which has recently been confirmed by Wagner (1961a). In Hulse's words: ". . . large as compared with small rewards produced greater resistance to extinction if partial reinforcement were used, but *less* resistance to extinction if continuous reinforcement were used . . ." (p. 56). In the cognitive language of Festinger this would have to mean that the subjects develop "some extra preference about something in the situation" (greater resistance to extinction) when rewards are *small and continuous;* and that they do not develop the extra preference, in fact quite the opposite, when rewards are *small and partial.* The result of Hulse and of Wagner is clearly in line with the S-R interpretation in terms of anticipatory responses: Extinction is *faster* following continuous rewards of large than of small magnitudes because of the greater FE in extinction and subsequent faster development of r_F. However, in partial reinforcement there is opportunity to elicit FE and to build r_F during acquisition. Under these circumstances, the larger magnitude of reward permits the mechanism for the PRE—the connection of s_F to continued approach—to develop more strongly. Consequently extinction should be *slower* following partial rewards of large magnitude than following partial rewards of small magnitude.

I am claiming that the S-R conditioning analysis of nonreward effects has much to recommend it. The question is, How much? To break the question down: Does inclusion of the concepts r_R and r_F in the analysis of partial reinforcement and discrimination learning provide us with explanatory power which a more parsimonious approach-avoidance analysis (e.g., D'Amato & Jagoda, 1960; Nissen, 1950) does not? Does this type of analysis of nonreward effects provide predictive power which a cognitive treatment of nonreward (e.g. Festinger, 1961) does not?

My attempts to convince the reader that the answer to these last two questions is "yes" will take two forms: (*a*) Some data will be presented which bear on predictions already made on

the basis of this type of analysis. I will claim that data such as these are not readily deducible from either a less specific cognitive language or from a more parsimonious approach-avoidance language. (*b*) I will offer an extension of frustrative nonreward theory designed to generate predictions about the course of discrimination learning following various prediscrimination experiences with partial and continuous reward.

Before proceeding to this extension of the theory of frustrative nonreward, a brief comment on the relation of the present position on nonreward to other theories of frustration. As I have indicated before (Amsel, 1958), ". . . we have been interested in certain active properties of nonreward following reward and no more than this is meant by the term frustration as we use it" (p. 103). Nevertheless, the mechanism which has been outlined to explain the PRE, particularly the conditioning of r_F and the connection of s_F to continued approach as a result of intermittent reward, is certainly a candidate for the learning theory counterpart of "frustration tolerance" as employed in social psychology and psychopathology. The treatment of frustration from the viewpoint of theories of personality is beyond the scope of this paper; and there is no intent here to grapple with the multiple meanings the concept has in those contexts. However, it would seem possible to recast portions of the frustration theory of Rosenzweig (e.g. 1944) which deal with nonreward (failure or delay of "gratification") into the present terminology; and the treatment of frustration by Rotter (1954), as a discrepancy between reward expectancy and its occurrence, seems compatible with the mechanisms hypothesized here and with the theoretical extension I will propose.

NONREWARD EFFECTS IN DISCRIMINATION LEARNING: SOME DATA

According to the sequence of five hypotheses on the role of nonreward and partial reinforcement in discrimination, it should be possible to show that the FE, measured in Runway 2 of a double runway, bears certain relationships to various stages of discrimination learning. More specifically, (*a*) evidence of nonreward related frustration, measured in Runway 2, should precede any evidence that the subjects are running faster to S+ and slower to S− in Runway 1, and (*b*) at some point following the appearance of discriminative behavior in Runway 1, a reduction in the magnitude of the FE measured in Runway 2 should become apparent.

In an unpublished doctoral dissertation, Ward (1961) employed the double runway to record discrimination to positive (e.g., black) and negative (e.g., white) cues presented singly in Runway 1, and the concurrent development of frustrative effects in Runway 2. A single runway condition (Runway 1 alone) was used to study the same discrimination uncontaminated by Runway 2 performance. There were 20 subjects in the double runway group and 10 subjects in the single runway comparison group. They were run six trials a day for 30 days, and the intertrial interval was about 10 minutes. Figure 1 presents data for the 20 subjects in the double runway. The top panel shows the developing FE in Runway 2. The bottom panel shows discrimination starting at about Day 7 (42 trials), and at a point well after the beginning of FE development on Day 3 (18 trials). There is an indication that the FE almost disappears (perhaps only temporarily) at a point where separa-

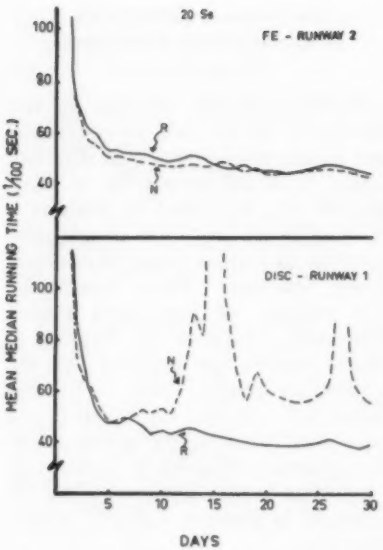

FIG. 1. The appearance and diminution of the frustration effect in Runway 2 (top panel) in sequential relationship to the development of the successive discrimination in Runway 1 (bottom panel). (From Ward, 1961)

tion of the discrimination curves is greatest. This occurs at about Day 15 (90 trials).

It could be argued the Ward result is an artifact: that different subjects reach criterion at different times. Anticipating this, I selected the data of seven of Ward's subjects which reached the discrimination criterion on the same block of trials. Figure 2 presents these data in 2-day blocks and shows essentially the same relationships as in Figure 1.

Figure 3 presents data from the single runway comparison group and shows the Runway 1 discrimination unaffected by any subsequent Runway 2 performance. Even in the absence of the second runway there is a tendency for the N and R curves of the discrimination to converge on later trials (see Berry, Verplanck, & Graham, 1943). These data of Ward, which have been duplicated in our laboratory as by-products of other studies, are taken as evidence for the involvement of frustrative effects in discrimination in a manner such as our analysis suggests.

In another experiment using the same double runway apparatus, Ward (1961) studied rate of discrimination learning in Runway 1 as related to number of prediscrimination rewards. In this experiment, which was similar to an earlier one of Shoemaker (1953) performed in a single runway, the response of running in Runway 1 and entering G_1 was rewarded 0, 12, or 48 times prior to learning a black-white discrimination in Runway 1. An implication of our position is that, with initial positive tendencies to the two discriminanda equal, rate of dis-

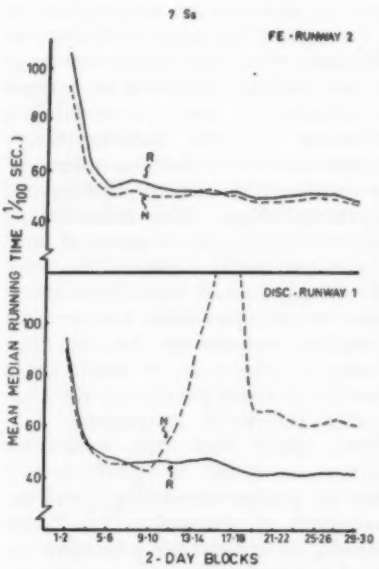

FIG. 2. These are the relationships shown in Figure 1 for seven subjects selected because they reached the criterion of discrimination on the same block of trials. (Adapted from Ward, 1961)

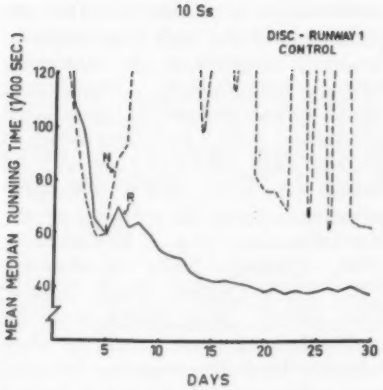

FIG. 3. The successive discrimination in a group which ran only in Runway 1. (Extreme N times going up to 3 or 4 seconds are not shown. Note the decline of the nonreward curve in the single runway group as in the double runway groups.)

crimination learning should be a positive function of the strength of r_R to the negative stimulus. This meant that prediscrimination trials were all rewarded in G_1 and that Runway 1 was black on half of these trials and white on the other half. On discrimination trials, Runway 1 was black when G_1 contained reward and white when G_1 contained no reward (or vice versa). Second runway performance was measured on every discrimination trial following reward or nonreward in G_1 to get a measure of the FE, whose strength at different stages could then be related to the developing discriminative performance. Table 1 shows the results of this experiment for both running time and goal entry measures for two criteria of learning. A more detailed analysis of the data in Ward's dissertation also indicates the appearance of the FE before discrimination and its diminution or disappearance following discrimination. These data are in accord with, and go a little beyond, previously cited studies which stress the importance of nonreward in discrimination.

A recent experiment by Jenkins (1961a) demonstrates the importance of nonreward effects in discrimination learning in another way. Using pigeons as subjects, a discrete trial procedure, and key pecking as the criterion response, Jenkins demonstrated that the positive stimulus elicits twice as many responses to extinction following discrimination training as does the same stimulus following an equal number of continuous reinforcements. On the basis of an ingenious arrangement of controls, Jenkins is able to conclude that the increased resistance to extinction is not due to the fact that the positive stimulus is eventually discriminated from the negative, but rather to the occurrence of nonreinforced responses in the process of learning the discrimination.

FRUSTRATIVE NONREWARD THEORY AND PREDISCRIMINATION EFFECTS

In the Ward experiment, subjects were exposed before a discrimination

TABLE 1

MEAN DAY ON WHICH CRITERION OF DISCRIMINATION REACHED BY GROUPS FOLLOWING 0, 12, OR 48 PREDISCRIMINATION B+ W+ TRIALS

Condition	Group			F
	0	12	48	
Running time				
Easy criterion	8.0	6.5	4.5	3.64*
Hard criterion	12.8	10.3	8.6	5.26**
Goal entry time				
Easy criterion	7.2	4.3	4.1	3.22*
Hard criterion	10.9	8.1	6.1	6.14**

* $p < .05$.
** $p < .01$.

to both of the stimuli, and approach responses in the presence of both were always rewarded. The Ward experiment represents continuous (prediscrimination) reward experience with both discriminanda. Now let us extend frustrative nonreward theory to a class of prediscrimination treatments broader than that represented by the Ward experiment.

Consider how discrimination learning will be affected not only by *continuous* prediscrimination reward experience but also by *partial* prediscrimination reward experience in relation to *one* or to *both* of a pair of discriminanda. I will restrict the discussion, in this paper, to a black-white, successive discrimination (single presentation of stimuli). The problem is to make predictions about the course of such a discrimination on the basis of some history of consistent or of partial reward experience in relation to one or both of the eventual discriminanda. Discrimination is measured in terms of amplitude (running time) changes to the positive and negative discriminanda.

Table 2 lists several possible kinds of exposure to discriminanda prior to a black-white discrimination. As we examine these cases, it is clear that theorists of learning have been interested mainly in prediscrimination experience involving both discriminanda. An early example is the continuity versus noncontinuity experiments which varied amount of presolution experience with the discrimination problem itself (B+W−) or with its reverse (B−W+), and with receptor orienting activity in relation to the discriminanda (e.g., Ehrenfreund, 1948; Gatling, 1951; Krechevsky, 1932, 1938; Lashley, 1942; Spence, 1940, 1945). More recent work on prediscrimination treatments as they influence later discrimination learning has taken several forms including Harlow's (e.g., 1949, 1950, 1957) experiments on learning sets, Lawrence's (1949, 1950, 1952) studies on acquired distinctiveness of cues, and the work relating overlearning to discrimination reversal and transfer (e.g., Birch, Ison, & Sperling, 1960; Capaldi & Stevenson, 1957; Pubols, 1956; Reid, 1953). These newer interests in prediscrimination experience have been somewhat less oriented than the earlier ones toward theoretical argument. In fact, they have to some extent tended to bring together the earlier wide split between S-R and cognitive approaches to discrimination, while investigating mediational processes affecting discrimination. Perhaps a point has been reached where it may be profitable to begin looking at the prediscrimination-discrimination sequence in a way which will bring the cognitive-expectancy position even more into the fold of an S-R (conditioning-expectancy) interpretation. It seems to me that a treatment of discrimination phenomena in terms of anticipatory (conditioned) mechanisms does just this.

TABLE 2

KINDS OF PREDISCRIMINATION EXPERIENCE IN RELATION TO A BLACK (+) WHITE (−) DISCRIMINATION

One Prediscrimination Stimulus
B+ Continuous Reward, Positive Stimulus
W+ Continuous Reward, Negative Stimulus
B± Partial Reward, Positive Stimulus
W± Partial Reward, Negative Stimulus
Two prediscrimination stimuli
B+ W+ Continuous Reward, Both Stimuli
B± W± Partial Reward, Both Stimuli
B− W+ Discrimination Reversal

Consider the following "attributes" of prediscrimination exposure to a single stimulus: (*a*) The prediscrimination stimulus may become either the positive or the negative stimulus in the eventual discrimination. (*b*) The re-

sponse to this prediscrimination stimulus may be partially or continuously rewarded. (*c*) The number of prediscrimination exposures to the stimulus-response-reward (or nonreward) sequence may be large, small, or of some intermediate value.[4] Figure 4

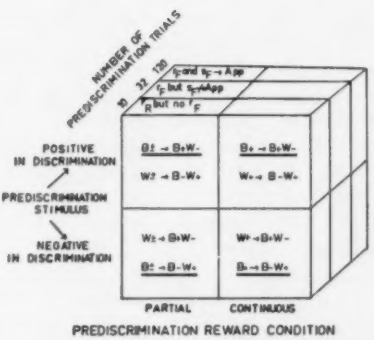

FIG. 4. Schema showing 24 prediscrimination-discrimination sequences (12 conditions, each counterbalanced) resulting from three attributes of prediscrimination exposure to a single discriminandum.

shows the various possible combinations of *a*, *b*, and *c* for a black-white discrimination. The 12 cells of this schema represent a matrix of possible experimental conditions, each counterbalanced for absolute effects of color.

What can be predicted from a behavior theory with frustrative nonreward premises about the effects of prediscrimination experiences in relation to a single discriminandum on subsequent black-white discrimination? Is it possible to show that all combinations of *a*, *b*, and *c* and also effects of prediscrimination experience involving both discriminanda will be deducible from a relatively small number of principles? In the next four sections, I will look at four kinds of prediscrimination treatment and, for each one, try to see what the theory leads us to expect.

Partial Reward of a Response to One Discriminandum Prior to Discrimination

Our guiding conception, when the prediscrimination treatment involves

[4] The values for number of trials, 10, 32, and 120, which appear in Figures 4, 5, and 6, and in the text in various places, are specified here for purposes of exposition and because we are now conducting experiments which involve these numbers of prediscrimination trials. However, these numbers of trials only represent guesses, at the present time, as to when the various stages of practice are reached. Rather than be pinned down to these numbers of trials in the predictions for discrimination learning which follow, I would prefer to set up independent indices of the existence of one stage or another and to predict discrimination outcome on the basis of these indices rather than some number of trials. The index for the first stage is relatively safe and simple. It is a stage at which nonreward produces no evidence of frustration as measured by the FE in a double runway, or in any other way. If there is no frustration, there can be no r_F. The presence of r_F takes us into the second stage. The choice of a number of trials to designate the second stage is most difficult since it must represent a very critical phase in partial reinforcement: enough trials for r_F to be present but not so many that s_F is associated with approach. The independent indicator of the middle stage (r_F, but s_F not associated with approach) is increased variability of behavior soon after nonreward produces the FE (Amsel, 1958). Stopping partial reward acquisition at such a point defines this middle stage. The late stage (identified here, conservatively, as 120 trials) has a clear independent definition. It exists (*a*) when variability decreases again, but more particularly (*b*) when partial reward performance exceeds continuous reward performance in vigor (see Goodrich, 1959; Haggard, 1959; Spence, 1960; Wagner, 1961a). While it might appear from earlier partial reinforcement experiments that 32 trials of partial reinforcement are sufficient to get us to the last stage and produce the PRE (and therefore are too many trials for our intermediate condition) experiments demonstrating PRE with this many acquisition trails (e.g., Sheffield, 1949; Weinstock, 1954) have usually involved an initial period of about 10 continuous rewards. Our procedure involves partial reward from the outset.

partial reinforcement and a sufficiently large number of trials, is that anticipatory frustration produced stimuli (s_F) become associated with approach responses. (Miller, 1960, has suggested a similar mechanism—"learning to advance to the cue of fear"—to account for the learning of resistance to fear.) The reasoning is that while s_F will at first elicit responses antagonistic to approach, it will finally become connected to approach tendencies. This means that at some stage in partial reinforcement s_F will become part of a stimulus complex evoking approach, and earlier evidence of conflict will disappear.

Of the four stages of practice in partial reinforcement previously outlined, three are theoretically differentiable with respect to the involvement of r_F, and therefore s_F: a stage where r_R has been conditioned, but not yet r_F (10 trials); a stage where both r_R and r_F have been conditioned and their stimuli, s_R and s_F, still evoke competing (approach and avoidance) response tendencies (32 trials); and a stage where both r_R and r_F are conditioned and the stimuli they produce, s_R and s_F, both evoke approach tendencies (120 trials). This conceptualization of three stages in partial reinforcement in regard to r_F—s_F, along with an assumption borrowed from Miller's analysis of conflict and displacement (1944, 1948), will provide the basis for some deductions which I find interesting.

The assumption from Miller's analysis of conflict, for which there is good experimental support, is that the generalization gradient for positive (approach) tendencies is flatter than that for negative (avoidance) tendencies. If we now identify $s_R \rightarrow$App (which results from r_R) and $s_F \rightarrow$Av (which results from r_F) as such positive and negative tendencies, the three panels of Figure 5, from top to bottom respectively, represent the state of affairs at the beginning of a discrimination, when the discrimination has been preceded by a few, an intermediate number, or a large number of partial reinforcements of a response to the prediscrimination stimulus (B). Represented on the left-hand side of the graph are the relative strengths of approach and avoidance tendencies to the prediscrimination stimulus (B) through the mediating response produced stimuli, s_R and s_F. Shown on the right are the generalized strengths of these tendencies to the stimulus (W) which *will be new* in the discrimination. (If the prediscrimination stimulus were W, then B would be the new stimulus in the discrimination and the B and W positions on the baseline would of course be reversed.)

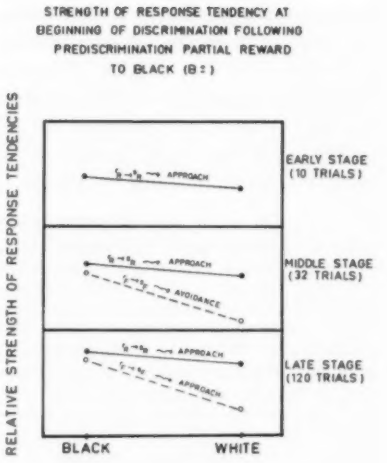

FIG. 5. An application of an assumption from Miller's analysis of conflict to predictions of ease of discrimination learning following various prediscrimination exposures to a single discriminandum.

A simplifying assumption, represented in the bottom panel of Figure 5, is that the strength of approach, elicited through r_F—s_F by B and W in the late

stage, is related to the strength of avoidance tendencies elicited through r_F—s_F in the middle stage. I am suggesting that the avoidance tendency to s_F is "neutralized" and has the effect of an approach tendency at this stage.

Some predictions can now be made relative to the conditions in Figure 4, and these will be restricted to the underlined conditions. (The other conditions simply counterbalance for any absolute effects of color and I assume they will not be different in direction.) Again, let me emphasize that the predictions will differ depending on the depth dimension: Was prediscrimination training discontinued at the stage of development of r_R but not r_F, indicated by the point of earliest appearance of FE? Or, was training carried on to the intermediate conflict stage at which s_F evokes avoidance? Or, were a very large number of trials run, allowing s_F to become connected to approach? Outlined below, for each of these cases, are predictions as to which discrimination will be learned faster: that in which the prediscrimination stimulus remains positive in the discrimination (B±→B+W−), or that in which it becomes negative (B±→B−W+).

Let us begin with the case of discrimination following 120 partial rewards. The reasoning (see Figure 5): At the start of discrimination, B will elicit r_R and r_F strongly and about equally. On the other hand, W, which is new in the discrimination, will elicit r_R relatively strongly because of the flat positive gradient, but will elicit r_F very weakly because of the steep negative gradient. This means that when B is made negative in the discrimination after a great many prediscrimination trials, B− will elicit strong r_F—s_F, and s_F will elicit continued approach. This will slow discrimination relative to the condition where B is made positive in the discrimination. Here the W− stimulus which is new in the discrimination elicits very weak r_F. Consequently, there is little s_F to elicit approach, little to counteract the build-up of avoidance to W−, and discrimination should develop more quickly. The theory predicts the following: After a very *large* number of prediscrimination partial rewards of an approach response to B, B+W− will produce *faster* discrimination than will B−W+. This prediction and its relationship to other predictions are shown in Figure 6.

The most interesting and critical comparison to be made for purposes of our analysis is between the 120-prediscrimination-trial case and the intermediate, 32-prediscrimination-trial case. In the 120-trial case s_F is already connected to approach, while in the 32-trial case s_F still elicits avoidance. Consequently the predictions in these two cases are reversed. The

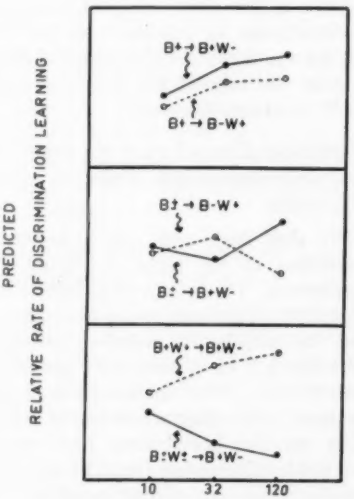

FIG. 6. Predicted rate of discrimination learning following various prediscrimination exposures to one or both of the eventual discriminanda.

reasoning for the intermediate case is that both r_R and r_F are being evoked by B at the start of discrimination, but s_R elicits approach and s_F elicits avoidance. When, after 32 trials, we switch from prediscrimination B± to the B+W− discrimination, B+ evokes both approach and avoidance tendencies through s_R and s_F, and W−, new in the discrimination, evokes mainly generalized approach (r_R) and little avoidance (r_F). When the shift is from B± to B−W+, there is a strong tendency for s_F to elicit avoidance in relation to the now negative B− (generalizing very little to W+) and a strong tendency for s_R to elicit approach to the new, now positive, W+. This latter situation is much more favorable for discrimination. The prediction: After an *intermediate* number of partial rewards of an approach response to B, B+W− will produce *slower* discrimination than will B−W+.

After a few B± trials, enough to produce some r_G but little or no FE and no r_F, there should be no difference in the rate of the B+W− and B−W+ discriminations.

Continuous Reward of a Response to One Discriminandum Prior to Discrimination

We deal here with the underlined conditions on the right of Figure 4: continuous (B+) prediscrimination experience followed by a discrimination in which B remains positive (B+W−) or becomes negative (B−W+). The expectations here are most interesting when we compare them to the predictions just made for prediscrimination partial reward. Briefly, with increasing numbers of rewards of approach responses to the prediscrimination stimulus, B+ should evoke increasing r_R, and nonreward in the subsequent discrimination should be increasingly frustrating. Since r_R generalizes strongly from B to W, both B and W should evoke r_R quite strongly and almost equally in the discrimination. Nevertheless, the FE should be somewhat greater when B is made the negative stimulus in the discrimination than when it stays positive. However, in either case (B+→B+W− or B+→B−W+), the flat gradient of generalization of positive tendencies leads to a prediction which is the same as the earlier Ward result: increasing the number of continuous prediscrimination rewards should produce faster discrimination.

In contrast to the center panel of Figure 6 in which the relationships at the middle and end points are reversed, the differences between the two curves in the top panel are shown as consistently in favor of faster discrimination for B+→B+W−. This is not a prediction based on frustrative nonreward factors alone. In fact, part of this is not a prediction at all. There is evidence in some unpublished data[5] that a switch from B+ in prediscrimination to B− in discrimination does produce a greater FE in Runway 2 than a switch from B+ to W−, as expected (Amsel & Hancock, 1957); however, an immediate generalization decrement of the Runway 1 response in switching from B to W weakens response to W, which speeds the apparent separation of the B+W− discrimination curves and retards the separation of the B−W+ curves. The suggestion is that, in the case of prediscrimination continuous reinforcement, the factor of generalization decrement, operating against the somewhat greater FE when the prediscrimination stimulus becomes negative in the discrimination, should maintain the relative advantage of B+→B+W− over B+→B−W+

[5] The reference is to data collected by Susan Hicks at Newcomb College, Tulane University, as part of an undergraduate honors project.

for all numbers of prediscrimination trials. In prediscrimination partial reinforcement (middle panel of Figure 6), since there is a switch-over in relative positions from 32 to 120 trials, the factor of generalization decrement should reduce the apparent difference at the intermediate point and increase the reversed difference at the 120-trial point.

While some of the implications of the analysis of variables represented by the schema in Figure 4 are currently being tested using a runway apparatus with amplitude changes as the response measures, we plan also to test these ideas in terms of rate of discrimination in two-choice situations with single presentation of discriminanda. In these latter situations, generalization decrement effects due to introduction of new stimuli would still be expected to effect running speed; however, they might have less effect on a correct-choice indicant of speed of discrimination, and provide an additional check on the implications of this position.

We have been looking at situations in which the response to a single stimulus cue is partially or continuously rewarded some number of times, and this stimulus then becomes either the positive or the negative cue in a discrimination. The last two sections of this paper will outline some of the factors operating in prediscrimination experience involving both discriminanda.

Partial versus Continuous Reward of Responses to Both Discriminanda Presented Separately Prior to Discrimination

The bottom panel of Figure 6 summarizes predictions about the rate of discrimination learning following prediscrimination experience with *both* discriminanda presented separately. I am suggesting that B+W+→B+W− operates essentially like B+→B+W− (or B+→B−W+), except that it would not be susceptible to generalization effects, since neither B nor W is new in the discrimination. There is some evidence related to this condition from the Shoemaker and Ward experiments, but only for the part of the "function" covering the first two points. Analysis of B±W± as a prediscrimination condition indicates a decreasing monotonic relationship between rate of discrimination and number of such prediscrimination trials. The reasoning is simply that partial reinforcement of approach responses to both stimuli will result in conflicting approach and avoidance response tendencies to *both* B and W at the 32-trial stage; and the elicitation of approach responses to s_F in connection with both B and W at the 120-trial stage. While the former represents a difficult base on which to build discrimination, as compared with B+W+, the latter should be even worse for discrimination, since it involves the mechanism for the PRE (s_F connected to approach) elicited by the negative stimulus in the discrimination. Consequently, a prolonged period of continued approach to both B+ and W− would be expected in the discrimination phase.

Discrimination Reversal as a "Prediscrimination" Phenomenon

The evidence on discrimination reversal is certainly not unequivocal; however, the weight of it suggests a U shaped relationship between ease of discrimination learning (B+W−) and number of prior B−W+ trials. This would look like the B±→B+W− prediction (middle panel Figure 6) in direction. Reversal, at least up to the 32-trial point, would produce successively slower learning; but reversal after many postcriterion trials, e.g., the 120-trial point, would facili-

tate the formation of the reversed discrimination.

The discrimination reversal studies cited earlier have been conducted in situations involving joint presentation of stimuli. They involve two basic findings. The first concerns ease of discrimination following the first reversal; the second concerns the "learning to learn" phenomenon—that successive reversals shorten the numbers of trials required to criterion. For the first of these findings, there is scant evidence in situations involving single presentation of stimuli (the present case); however the recent Birch, Ison, and Sperling study (1960) is important and relevant.

The conditions of the Birch, Ison, and Sperling experiment are closest to meeting the boundary conditions of a frustrative nonreward analysis of reversal as a "prediscrimination" phenomenon. Their subjects, rats, were given 11 trials each day, the first always positive, the next 10 divided between positive and negative according to random orders. Their measure was response speed rather than choice, and their apparatus was essentially a runway. They were trained initially to approach a white platform and to avoid a black one and, when they reached a criterion of discrimination defined in terms of approach speed, half of the subjects were switched to a B+W− discrimination immediately, the other half after 12 additional days of overtraining to B−W+. The average number of trials to reach the original criterion for both groups was about 50 (well beyond our 32-trial point). This means that the overtrained group had about 180 trials to B−W+ (well beyond our 120-trial point) before it was reversed and learned B+W−. The overtrained group learned the reversed discrimination in significantly fewer trials.

Interpreting their results, Birch, et al. (1960) state: ". . . the difference between the two groups appears to be located in their differential rates of extinction to the previously reinforced cue. No difference in rate of acquisition to the newly positive cue was found in the present data" (p. 39). These results have a rather obvious interpretation in terms of a frustrative nonreward position; but the situation is complex and there are theoretical difficulties to overcome. The obvious part of the interpretation is that the overlearned group has developed stronger r_R to W+, which makes subsequent W− more frustrating, which facilitates the conditioning of W to r_F, which, finally, ensures the evocation of avoidance responses by s_F. This part is fine for the theory; however, there is the matter of what happens to s_F, which presumably develops in the initial discrimination of the Birch, et al. experiment to the negative stimulus (B−) and which is connected to avoidance. Shouldn't the overtrained group have a stronger $s_F \rightarrow$ avoidance tendency in relation to B than the criterion group? And shouldn't this retard the learning of the reversed discrimination in the overtrained group when B subsequently becomes the positive stimulus? This difficulty remains for the theory unless one assumes that r_R (and $s_R \rightarrow$ approach) builds in strength very slowly over many trials (at least over 50) while r_F (and $s_F \rightarrow$avoidance) develops quickly and is maximal by about 50 trials—this being the case despite the fact that the growth of r_F depends upon the prior growth of r_R. This is certainly an interesting possibility.

SUMMARY

This paper begins with an account of recent evidence for a concept of frustrative nonreward in behavior

theory, and presents support for the position that an "active" role must be assigned to nonreward in both motivation and inhibition. This evidence is not only from experiments specifically designed to study frustrative properties of nonreward, but also from a variety of other kinds of experiments, including, among others, studies of extinction as a function of magnitude of reward and number of rewards, and experiments in discrimination learning which demonstrate the relative importance of nonrewarded and rewarded trials.

The importance of an active conception of nonreward in current thinking is discussed with reference to similarities and differences between recent cognitive-expectancy and neo-Hullian interpretations of the partial reinforcement effect. The suggestion is that current S-R conditioning interpretations of these partial reinforcement phenomena, stressing the role of both anticipatory reward and anticipatory frustration mechanisms, and involving a more specific and better worked out language, provide a more workable "expectancy" theory than do cognitive interpretations.

The contention is that current neo-Hullian approaches, employing conditioning-expectancy concepts, go beyond cognitive-expectancy approaches in both specificity and predictive power. As evidence for the merit of such a position, data are presented from discrimination learning experiments which test predictions from a conditioning analysis of nonreward effects, and an extension of frustrative nonreward theory is presented which leads to predictions about the course of discrimination learning following various prediscrimination experiences. These predictions depend on the interaction of three attributes of prediscrimination experience: (*a*) whether prediscrimination exposure is to the stimulus which is eventually positive or negative in the discrimination (or to both); (*b*) whether the prediscrimination treatment involves continuous or partial reward; and (*c*) the number of prediscrimination exposures (trials), representing different stages of development of anticipatory reward and anticipatory frustration.

REFERENCES

ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

ADELMAN, H. M., & ROSENBAUM, G. Extinction of instrumental behavior as a function of frustration at various distances from the goal. *J. exp. Psychol.*, 1954, **47**, 429–432.

AMSEL, A. A three-factor theory of inhibition: An addition to Hull's two-factor theory. Paper read at Southern Society for Philosophy and Psychology, Roanoke, 1951.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

AMSEL, A., ERNHART, C. B., & GALBRECHT, C. R. Magnitude of frustration effect and strength of antedating goal factors. *Psychol. Rep.*, 1961, **8**, 183–186.

AMSEL, A., & HANCOCK, W. Motivational properties of frustration: III. Relation of frustration effect to antedating goal factors. *J. exp. Psychol.*, 1957, **53**, 126–131.

AMSEL, A., & PENICK, E. The influence of early experience on the frustration effect. *J. exp. Psychol.*, 1962, **63**, 167–176.

AMSEL, A., & PROUTY, D. L. Frustrative factors in selective learning with reward and nonreward as discriminanda. *J. exp. Psychol.*, 1959, **57**, 224–230.

AMSEL, A., & ROUSSEL, J. Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *J. exp. Psychol.*, 1952, **43**, 363–368.

AMSEL, A., & WARD, J. S. Motivational properties of frustration: II. Frustration drive stimulus and frustration reduction in selective learning. *J. exp. Psychol.*, 1954, **48**, 37–47.

ARMUS, H. L. Effect of magnitude of reinforcement on acquisition and extinction

of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

BERRY, R. N., VERPLANCK, W. S., & GRAHAM, C. H. The reversal of discrimination in a simple running habit. *J. exp. Psychol.*, 1943, **32**, 325–334.

BIRCH, D. Discrimination learning as a function of the ratio of nonreinforced to reinforced trials. *J. comp. physiol. Psychol.*, 1955, **66**, 456–464.

BIRCH, D., ISON, J. R., & SPERLING, S. E. Reversal learning under single stimulus presentation. *J. exp. Psychol.*, 1960, **60**, 36–40.

BITTERMAN, M. E., FEDDERSON, W. E., & TYLER, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

BOWER, G. H. A contrast effect in differential conditioning. *J. exp. Psychol.*, 1961, **62**, 196–199.

BOWER, G. H., & STOCKS, D. Graded frustration effect as a function of graded reduction in reward. Paper read at Psychonomic Society, Chicago, 1960.

BROWN, J. S., & FARBER, I. E. Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration. *Psychol. Bull.*, 1951, **48**, 465–495.

BUSH, R. R., & MOSTELLER, F. A mathematical model for simple learning. *Psychol. Rev.*, 1951, **58**, 313–323.

CANTOR, G. N., & SPIKER, C. C. Effects of nonreinforced trials on discrimination learning in preschool children. *J. exp. Psychol.*, 1954, **47**, 256–258.

CAPALDI, E. J., & STEVENSON, H. W. Response reversal following different amounts of training. *J. comp. physiol. Psychol.*, 1957, **50**, 195–198.

D'AMATO, M. R., & JAGODA, H. Effects of extinction trials on discrimination reversal. *J. exp. Psychol.*, 1960, **59**, 254–260.

DENNY, M. R., & DUNHAM, M. D. The effect of differential nonreinforcement of the incorrect response on the learning of the correct response in the simple T-maze. *J. exp. Psychol.*, 1951, **41**, 382–389.

EHRENFREUND, D. An experimental test of the continuity theory of discrimination learning with pattern vision. *J. comp. physiol. Psychol.*, 1948, **41**, 408–422.

ENINGER, M. U. The role of generalized approach and avoidance tendencies in brightness discrimination. *J. comp. physiol. Psychol.*, 1953, **46**, 398–402.

ESTES, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

ESTES, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of science.* Vol. 2. New York: McGraw-Hill, 1959.

FERSTER, C. B. Withdrawal of positive reinforcement as punishment. *Science*, 1957, **126**, 509.

FERSTER, C. B. Control of behavior in chimpanzees and pigeons by time out from positive reinforcement. *Psychol. Monogr.*, 1958, **72**(8, Whole No. 461).

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

FESTINGER, L. The psychological effects of insufficient rewards. *Amer. Psychologist*, 1961, **16**, 1–11.

FITZWATER, M. E. The relative effect of reinforcement and nonreinforcement in establishing a form discrimination. *J. comp. physiol. Psychol.*, 1952, **45**, 476–481.

GATLING, F. A study of the continuity of the learning process as measured by habit reversal in the rat. *J. comp. physiol. Psychol.*, 1951, **44**, 78–83.

GOODRICH, K. P. Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, 1959, **57**, 57–63.

GRICE, G. R., & GOLDMAN, H. M. Generalized extinction and secondary reinforcement in visual discrimination learning with delayed reward. *J. exp. Psychol.*, 1955, **50**, 197–200.

GROVE, G. R., & ENINGER, M. U. The relative importance of approach and avoidance tendencies in brightness discrimination learning. Paper read at Midwestern Psychological Association, Chicago, 1952.

HAGGARD, D. F. Acquisition of a simple running response as a function of partial and continuous schedules of reinforcement. *Psychol. Rec.*, 1959, **9**, 11–18.

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARLOW, H. F. Analysis of discrimination learning by monkeys. *J. exp. Psychol.*, 1950, **40**, 26–39.

HARLOW, H. F., & HICKS, L. H. Discrimination learning theory: Uniprocess vs. duoprocess. *Psychol. Rev.*, 1957, **64**, 104–109.

HOLDER, W. B., MARX, M. H., HOLDER, E. E., & COLLIER, G. Response strength as a function of delay of reward in a runway. *J. exp. Psychol.*, 1957, **53**, 316–323.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HULL, C. L., LIVINGSTON, J. R., ROUSE, R. O., & BARKER, A. N. True, sham, and esophageal feeding as reinforcements. *J. comp. physiol. Psychol.*, 1951, **44**, 236–245.

HULSE, S. H., JR. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

HUMPHREYS, L. G. The effect of random alternation of reinforcements on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.

HUMPHREYS, L. G. Extinction of conditioned psycho-galvanic response following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–76.

JENKINS, H. M. The effect of discrimination training on extinction. *J. exp. Psychol.*, 1961, **61**, 111–121. (a)

JENKINS, H. M. Resistance to extinction when partial reinforcement is followed by various amounts of regular reinforcement. Paper read at Psychonomic Society, New York, 1961. (b)

KENDLER, H. H., PLISKOFF, S. S., D'AMATO, M. R., & KATZ, S. Nonreinforcements versus reinforcements as variables in the partial reinforcement effect. *J. exp. Psychol.*, 1957, **53**, 269–276.

KRECHEVSKY, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, **39**, 516–532.

KRECHEVSKY, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, **45**, 107–133.

LACHMAN, R. The influence of thirst and schedules of reinforcement-nonreinforcement ratios upon brightness discrimination. *J. exp. Psychol.*, 1961, **62**, 80–87.

LAMBERT, W. W., & SOLOMON, R. L. Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, 1952, **45**, 269–279.

LASHLEY, K. S. An examination of the "continuity theory" as applied to discrimination learning. *J. gen. Psychol.*, 1942, **26**, 241–265.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39**, 770–784.

LAWRENCE, D. H. Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40**, 175–188.

LAWRENCE, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.

LAWRENCE, D. H. Learning. *Annu. Rev. Psychol.*, 1958, **9**, 157–188.

LAWSON, R., & MARX, M. H. Frustration: Theory and experiment. *Genet. psychol. Monogr.*, 1958, **57**, 393–464.

LOGAN, F. A., BEIER, E. M., & KINCAID, W. D. Extinction following partial and varied reinforcement. *J. exp. Psychol.*, 1956, **52**, 65–70.

LONGSTRETH, L. E. The relationship between expectations and frustrations in children. *Child Develpm.*, 1960, **31**, 667–671.

MARZOCCO, F. M. Frustration effect as a function of drive level, habit strength and distribution of trials during extinction. Unpublished doctoral dissertation, State University of Iowa, 1951.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MILLER, N. E. Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context. *J. exp. Psychol.*, 1960, **60**, 137–145.

MILLER, N. E., & STEVENSON, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.

MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.

NISSEN, H. W. Description of learned responses in discrimination behavior. *Psychol. Rev.*, 1950, **57**, 121–131.

NORTH, A. J., & STIMMEL, D. T. Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, 1960, **6**, 227–234.

OLDS, J. The influence of practice on the strength of secondary approach drives. *J. exp. Psychol.*, 1953, **46**, 232–236.

OLDS, J. *The growth and structure of motives*. Glencoe, Ill.: Free Press, 1956.

PAVLIK, W. B. The effects of deprivation schedule and reward magnitude on performance during acquisition and extinction. Paper read at Eastern Psychological Association, Philadelphia, April 1961.

PENNEY, R. K. The effects of non-reinforcement on response strength as a function of number of previous reinforcements. *Canad. J. Psychol.*, 1960, **14**, 204–215.

PUBOLS, B. H., JR. The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, 1956, **49**, 243–248.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, **46**, 107–112.

REYNOLDS, W. F., RICHTER, M. L., & CARLOCK, J. Resistance to extinction as a function of reward magnitude and number of training trials. Paper read at Psychonomic Society, Chicago, 1960.

REYNOLDS, W. F., & SIEGEL, M. H. Drive reversal effect as a function of magnitude of reward. *Psychol. Rep.*, 1961, **8**, 135–141.

ROHRER, J. H. A motivational state resulting from non-reward. *J. comp. Psychol.*, 1949, **42**, 476–485.

ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.

ROTTER, J. B. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

ROUSELL, J. Frustration effects as a function of repeated non-reinforcements and as a function of the consistency of reinforcement prior to introduction of nonreinforcement. Unpublished master's thesis, Tulane University, 1952.

SEWARD, J. P. Experimental evidence for the motivating function of reward. *Psychol. Bull.*, 1951, **58**, 130–149.

SEWARD, J. P., PEREBOOM, A. C., BUTLER, B., & JONES, R. B. The role of prefeeding in apparent frustration effect. *J. exp. Psychol.*, 1957, **54**, 445–450.

SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.

SHEFFIELD, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.

SHOEMAKER, H. A. The relative efficiency of rewarded and non-rewarded training in a black-white discrimination problem. Unpublished doctoral dissertation, University of Colorado, 1953.

SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.

SOLOMON, R. L. The externalization of hunger and frustration drive. *J. comp. physiol. Psychol.*, 1956, **49**, 145–148.

SPENCE, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

SPENCE, K. W. Analysis of the formation of visual discrimination habits in the chimpanzee. *J. comp. Psychol.*, 1937, **23**, 77–100.

SPENCE, K. W. Continuous versus noncontinuous interpretations of discrimination learning. *Psychol. Rev.*, 1940, **47**, 271–288.

SPENCE, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. *Behavioral theory and learning.* Englewood Cliffs: Prentice-Hall, 1960.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.

THEIOS, J. The partial reinforcement effect sustained through blocks of continuous reinforcement. *J. exp. Psychol.*, 1962, in press.

TYLER, D. W., MARX, M. H., & COLLIER, G. Frustration stimuli in discrimination. *J. exp. Psychol.*, 1959, **58**, 295–301.

WAGNER, A. R. The treatment of inhibition in the more complex behavioral situation. Paper read at American Psychological Association, 1961. (b)

WAGNER, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *J. exp. Psychol.*, 1959, **57**, 130–136.

WAGNER, A. R. Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, 1961, **62**, 234–242. (a)

WARD, J. S. An experimental investigation of the role of frustrative factors in discrimination learning. Unpublished doctoral dissertation, Tulane University, 1962.

WEINSTOCK, S. Resistance to extinction of a running response following partial reinforcement under widely spaced trials. *J. comp. physiol. Psychol.*, 1954, **47**, 318–322.

WILSON, W., WEISS, E. J., & AMSEL, A. Two tests of the Sheffield hypothesis concerning resistance to extinction, partial reinforcement, and distribution of practice. *J. exp. Psychol.*, 1955, **50**, 51–60.

(Received November 20, 1961)

Psychological Review
1962, Vol. 69, No. 4, 329-343

THE SELECTION OF STRATEGIES IN CUE LEARNING [1]

FRANK RESTLE

Michigan State University

In a cue learning problem (discrimination learning, concept formation, and maze learning) the subject chooses one of two or more responses on each trial. The correctness of the response depends on some aspect of the situation at the time of response.

In this paper it is assumed that subjects have difficulty with cue learning problems to the degree that they tend to use strategies (habits, patterns of response) which conflict with the strategy intended by the experimenter. It is often assumed that the subject must also associate correct responses with cues in the situation. However, in most cue learning experiments the subject is instructed or pretrained to make the desired responses before the cue learning process begins. Furthermore, the experimental situation constrains the subject to make exactly one of the available responses per trial. It seems reasonable to assume that cue learning is not so much a matter of the formation as of the selection of responses. The present theoretical discussion begins with a set of strategies and is concerned with the mechanisms by which the subject might select out those strategies intended, and consistently rewarded, by the experimenter.

The term "strategy" is employed in a sense related to the more common terms, "habit" and "hypothesis," to designate a particular pattern of responses to stimuli. Consider, for example, a rat being trained to jump to the larger of two white circles in a Lashley Jumping Stand. Some possible strategies include; jumping to the left side, jumping to the right side, jumping alternately left and right, jumping always to the side last reinforced, jumping to the side which last was punished, jumping in the direction the rat was placed down on the stand, jumping toward the circle with less (or more) acute curvature, jumping toward the side with more white area, jumping toward the side which reflects more total light, and so forth. Each such pattern of behavior, if it occurs in isolation for a sufficiently long sequence of trials, can be identified by the experimenter. It will be seen that there are ordinarily a large number and variety of possible strategies in a cue learning problem, and that a number of strategies may be perfectly confounded. In the example above, jumping to the circle with less acute curvature will always lead to the same result as jumping to the side which reflects more light, since these will both be jumps to the larger circle.

Three Models of the Selection of Strategies

The general idea developed here is that the problem gives rise to a set of strategies. The subject uses these various strategies, which at first are selected at random, and attempts to select strategies which will consistently lead to correct responses. For simplicity the discussion is limited to

[1] This research was facilitated by the author's tenure as Faculty Research Fellow of the Social Science Research Council, 1959-61.

problems in which there are at least some strategies which are always correct.

Let H be the set of strategies available to the subject at the beginning of the problem. Suppose that some subset C of these strategies are always correct, another subset W are always wrong, and the remainder, I, are sometimes correct and sometimes wrong. For simplicity of presentation, begin with a problem in which the strategies of I are correct and wrong, on successive trials, with independent probabilities all of which are $\frac{1}{2}$, and restrict attention to the two-choice problem.

Presentation of the theory of this conventional discrimination learning problem will begin with a special case in which one strategy is used on each trial. Then an alternative theory will be proposed, in which the subject uses all strategies at once and attempts to narrow down to the correct one. Third, a model will be proposed in which the subject chooses a random sample of all strategies and attempts to narrow down to the correct strategies within his sample. These three models will be shown to be essentially equivalent, in a theorem which may be called the "indifference to sample size" theorem. Following proof of that theorem, some empirical implications of the theory will be derived and compared with relevant data.

One Strategy at a Time

On the first trial of the experiment, the subject selects a certain single strategy from H at random and makes the indicated response. If the response is correct, the same strategy is used on Trial 2. If it is again correct, it is used again on Trial 3, and so forth. If, on any trial, the response is incorrect, the subject returns his strategy to the set H and then chooses a strategy again at random. It is considered possible that the same strategy may be resampled, so that the sampling process is with replacement. Imagine that an error occurs on Trial 2. Then a new strategy is chosen and this strategy is used on Trial 3. If it is correct it is used again on Trial 4. If the response on Trial 3 is wrong, the subject again chooses an hypothesis at random for Trial 4.

Notice that if the subject chooses a correct strategy, he will make no more errors. The correct strategy leads to correct responses, and when the response is correct the subject does not change strategy. Thus, the learning process is terminated by the choice of a correct strategy.

Since sampling is with replacement, the probabilities of choosing a correct, a wrong, or an irrelevant strategy are constant from trial to trial. Let these three probabilities be c, w, and $i = 1 - c - w$.

At the beginning of training the subject chooses a strategy of one of the three types, which have probabilities c, w, and i. After each error he also chooses a strategy, with probabilities c, w, and i. Thus an error is an event which returns the subject to exactly the same condition (in terms of the probabilities of future events) with which he started—it resets the process and makes it begin over. Such an event is called a "recurrent event" (Feller, 1950). Since our experimental interest centers around errors and correct responses, we can limit attention to errors and thus to the theory of recurrent events.

Imagine that at some Trial n the subject makes an error. There is a certain probability f_1 that the next error occurs at Trial $n + 1$, a probability f_2 that the next error occurs at Trial $n + 2$, and in general a prob-

ability distribution f_j that the next error after Trial n occurs at Trial $n+j$. In a model of recurrent events like the present one, the distribution f_j is the same for all Trials n on which an error occurs, and is the same as the distribution of probabilities that the first error occurs on Trial 1, 2, etc.

The random process and all of its properties are specified by the distribution f_j. However, the analysis of the process from the distribution f_j will be postponed until two steps have been completed; first, calculation of the distribution f_j for the one-strategy-at-a-time model, and second, the formulation of two other models, very different in assumptions, which also lead to systems of recurrent events with the same distribution. When these steps are completed, the generality of the f distribution will be apparent, and derivations of properties of the process will be justified.

First, f_1 is the probability of an error immediately following another error. The second error can occur because a wrong strategy was chosen (with probability w) or because an irrelevant strategy was chosen and turned out to be wrong (with probability $\frac{1}{2}i$). These are mutually exclusive events and exhaust the possibilities, so that:

$$f_1 = w + \tfrac{1}{2}i.$$

An error occuring on the second trial after the last one, with a correct response intervening, cannot result from the selection of a wrong strategy for the wrong strategy would not lead to an intervening correct response. If the error does occur on the second trial we know that the strategy chosen was not correct. Hence it must be an irrelevant one which led first to a correct and then to a wrong response. In simple discrimination learning, the probability that an irrelevant strategy would lead to a correct and then a wrong response is $\frac{1}{2}\cdot\frac{1}{2} = (\frac{1}{2})^2$. The probability of choosing an irrelevant strategy which gives this sequence, correct-wrong, is therefore:

$$f_2 = i(\tfrac{1}{2})^2$$

By the same reasoning, any value of f_j, $j > 2$, implies that an irrelevant strategy was chosen and then led to $j-1$ correct responses followed by a wrong response. The probability of a string of exactly $j-1$ correct responses each with probability $\frac{1}{2}$, followed by an error with probability $\frac{1}{2}$, is $(\frac{1}{2})^j$. Hence, for $j \geq 2$:

$$f_j = i(\tfrac{1}{2})^j$$

Summarizing the results we have the distribution:

$$f_1 = w + \tfrac{1}{2}i$$
$$f_j = i(\tfrac{1}{2})^j$$

for all $j \geq 2$. The distribution f_j is not a proper probability function because it does not sum to unity. Notice that:

$$\sum_{j=1}^{\infty} (f_j) = w + i \sum_{j=1}^{\infty} (\tfrac{1}{2})^j = w + i = 1 - c$$

In Feller's terms this means that errors are *uncertain* recurrent events, for with probability c the subject chooses a correct strategy and never makes another error. Some of the probability of the f distribution is located at the (improper) point ∞, and the proportion so located is c. This merely reflects the fact that learning occurs in the present model and the subject can eliminate errors. The fact that f_j is not a probability function does not place any serious difficulties in the way of analysis.

All Strategies at Once

We now consider a second model of a subject working on the same dis-

crimination learning problem discussed above, with exactly the same set of strategies H divided into correct (C), wrong (W), and irrelevant (I) subsets. Since a strategy corresponds to any specific cue in the situation and these cues are numerous and finely divisible, consider that the set of strategies consists of a fairly sizeable (though finite) number of strategies, each of which is as likely to be used, and has as much influence on behavior, as every other.

In the all-strategies-at-once model we imagine that the subject begins the learning process by considering the entire set H of strategies simultaneously. In informal terms, he then attempts to narrow down to the correct strategy by a process of elimination. However, we suppose that the subject has only a limited memory and cannot remember all of the strategies at once, but only those with which he is presently concerned. The main disability of the subject in this model is that he may accidentally eliminate the correct strategies. When this happens he is unable to remember what they were, for he can remember only the ones he is still using. When the correct strategies have been lost the subject has no recourse but to begin over with the whole set of strategies. This general idea is now stated more precisely.

On the first trial the subject considers all strategies and chooses his response. The correct strategies and some of the irrelevant ones lead toward a correct response, the wrong strategies and the remainder of the irrelevant ones lead to an error. We suppose that the probability of a response is equal to the proportion of the strategies which lead to that response, so that, if we let N_c, N_w, and N_i be the numbers of strategies in sets C, W, and I, and if we let N_i^- be the number of irrelevant strategies which lead to an error on the next trial, the probability of an error is:

$$f_1 = (N_w + N_i^-)/(N_w + N_c + N_i).$$

When the subject makes this first response we suppose that he eliminates or sets aside all the strategies he does not use. Suppose, for example, that he chooses a correct response. He at the same time eliminates all the wrong strategies and the irrelevant strategies which would have led him to the opposite response. On the next trial he will have only the correct strategies and the surviving irrelevant strategies left.

To complete the discussion and generalize it somewhat we introduce notation for the set of irrelevant strategies which, following any Trial n, first lead to a wrong response on Trial $n + j$. Let $N_{n,j}$ be the number of irrelevant strategies which lead to correct responses on trials $n + 1$, $n + 2$, . . . , $n + j - 1$, and an error on trial $n + j$. Now we can write the probability of an error after an error as:

$$f_1 = (N_w + N_{n,1})/(N_c + N_w + N_i).$$

After one correct response, the only strategies which will lead to an error are those $N_{n,2}$ irrelevent strategies which first lead to an error on the second trial. The total strategies left are N_c correct and $N_i - N_{n,1}$ irrelevant strategies which have not already led toward a wrong response. Thus, the probability of an error following one correct response which follows a given error is:

$$f_2 = (1 - f_1)(N_{n,2}/[N_c + N_i - N_{n,1}])$$
$$= \frac{N_c + N_i - N_{n,1}}{N_c + N_w + N_i} \cdot \frac{N_{n,2}}{N_c + N_i - N_{n,1}}$$

Extension of this line of argument shows that in general:

$$f_j = \frac{N_c + N_i - N_{n,1}}{N_c + N_w + N_i} \cdot \frac{N_c + N_i - N_{n,1} - N_{n,2}}{N_c + N_i - N_{n,1}} \cdot \ldots$$
$$\cdot \frac{N_c + N_i - N_{n,1} - \ldots - N_{n,j-2} - N_{n,j-1}}{N_c + N_i - N_{n,1} - \ldots - N_{n,j-3} - N_{n,j-2}} \cdot \frac{N_{n,j}}{N_c + N_i - N_{n,1} - \ldots - N_{n,j-1} - N_{j-1}}$$

In this expression the numerator of each term except the last cancels with the denominator of the following term, so that:

$$f_j = N_{n,j}/(N_c + N_w + N_i)$$

In our ideal experiment, $N_{n,j}$, the number of strategies which first lead to a wrong response on Trial $n + j$ after a given error on Trial n, is independent of n and is given by $N_i(\frac{1}{2})^j$. This statement follows from the idea that irrelevant strategies are correct and wrong with independent probabilities of $\frac{1}{2}$. Actually, $N_i(\frac{1}{2})^j$ is the expected number of irrelevant strategies which first lead to an error on Trial $n + j$, but the derivation to follow does not hinge on the distinction.

If we substitute $N_i(\frac{1}{2})^j$ for $N_{n,j}$ in the expression for f_j and simplify, letting i be the proportion of irrelevant strategies in H, we obtain:

$$f_j = i(\tfrac{1}{2})^j.$$

A similar substitution in the expression for f_1 gives:

$$f_1 = w + \tfrac{1}{2}i.$$

These two results are in agreement with the one-strategy-at-a-time model. We have shown that the one-strategy-at-a-time and the all-strategies-at-once models are both systems in which errors are recurrent events, and we have shown that f_j is the same in both systems. Hence, so far as errors are concerned, the two models are identical.

One remark is in order. In the one-strategy-at-a-time model, the probability of a correct response on any trial is 1, $\frac{1}{2}$, or 0 depending on whether a correct, irrelevant, or wrong strategy is being used. The probability of a correct response in the all-strategies-at-once model can take a variety of values; if there are a great many strategies, the possible values of this probability are numerous and closely packed. The two models are not at all alike in the probabilities involved, and the theorem that they give indistinguishable sequences of errors is by no means obvious on intuitive grounds, even though the proof is easy.

The correspondence of the models extends to the interpretation of the parameters. In the one-strategy-at-a-time model the parameters c, w, and i are the probabilities of selecting a correct, wrong, or irrelevant strategy. In the all-strategies-at-once model the parameters c, w, and i are the proportions of correct, wrong, and irrelevant strategies, where elementary strategies are taken to have equal weight. In terms of experimental realizations of the models, these two sets of parameters are not distinguishable.

It is still possible, of course, to tell which model is more reasonable by any technique which goes beyond the choice data. One might, for example, ask human subjects which strategies they are using, or one might use memory tests to see what parts of the situation have been employed. The equivalence of the models is asserted only for the acquisition of a single problem, and only with respect to the overt choices made: the writer will suggest, in subsequent work, that rec-

ognition and transfer tests may permit differential predictions.

A Random Sample of Strategies

The two models discussed above are extremes. One would intuitively imagine that the subject would use, not all strategies and not just one, but some sample of strategies. The third model, called the random-sample-of-strategies model, supposes that on the first trial and after each error the subject draws a random sample of all strategies. So long as he makes no errors the subject discards wrong and irrelevant strategies by the same mechanism as that invoked in the all-strategies-at-once model. If the subject makes an error he takes a new independent sample from H and begins the process over.

The theorem to be proved is, if the sampling process is random in a strict sense to be described below and excludes the possibility that the subject might choose the empty sample, then the random-sample-of-strategies model generates a system of recurrent events which is the same as the other two models given above.

We again consider the set H of strategies, composed of equally potent elementary strategies. Suppose that there are N strategies in H of which N_c are correct, N_w are wrong, and N_i are irrelevant. We shall interpret N_c/N as c, the proportion of correct strategies, N_w/N as w, and N_i/N as i.

In a random sampling model of learning one may assign a fixed probability of being sampled to each element and let the sample size vary randomly, or assign a fixed sample-size, supposing that all such samples are equally probable (Estes, 1959.) To eliminate the possibility of the empty sample it is desirable to deal with a fixed sample size n. The random-sampling specification is that all samples of size n are equally probable.

The total number of samples of size n which can be drawn from a population of N elements is:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Similarly, one can draw $\binom{N_c}{n_c}$ samples of n_c correct strategies from the population of N_c strategies. For each of these samples of n_c correct strategies, there are $\binom{N_w}{n_w}$ samples of n_w wrong strategies, so that there are $\binom{N_c}{n_c}\binom{N_w}{n_w}$ ways of drawing n_c correct and n_w wrong strategies. For each of these combinations there are $\binom{N_i}{n_i}$ ways of drawing n_i irrelevant strategies. Hence the total number of ways one could obtain n_c correct, n_w wrong, and n_i irrelevant strategies is:

$$\binom{N_c}{n_c}\binom{N_w}{n_w}\binom{N_i}{n_i}$$

where, of course, it is assumed that $n_c + n_w + n_i = n$. Since every sample of size n is assumed to be as likely as any other, the probability of drawing a sample of exactly n_c correct, n_w wrong, and n_i irrelevant strategies is:

$$P(n_c, n_w, n_i) = \frac{\binom{N_c}{n_c}\binom{N_w}{n_w}\binom{N_i}{n_i}}{\binom{N}{n}}$$

which is the three-category generalization of the hypergeometric distribution (Feller, 1950).

It was asserted, in the statement of the model, that once the subject has chosen a sample his performance follows the rules of the all-strategies-at-once model. Hence, after any

error and the choice of a sample with n_c, n_w, and n_i strategies, we can construct the f_j distribution. We may write the f distribution conditional on the constitution of the sample as:

$$f(j|\text{sample with } n_c, n_w, n_i \text{ strategies}) = \begin{cases} (n_w + \frac{1}{2}n_i)/n & \text{for } j = 1 \\ (\frac{1}{2})^j n_i/n & \text{for } j > 1 \end{cases}$$

directly from our results on the all-strategies-at-once model. To obtain the unconditional f_j we multiply each conditional f by the probability of the sample, and sum. This gives us the expectation of the conditional f. For f_1 we have:

$$f_1 = \mathbf{E}[(n_w + \tfrac{1}{2}n_i)/n] = \mathbf{E}(n_w/n) + \tfrac{1}{2}\mathrm{E}(n_i/n)$$

However, in the hypergeometric distribution the mean proportion of wrong strategies in samples of size n is just the proportion of wrong strategies in the whole population. That is, $\mathbf{E}(n_w/n) = N_w/N = w$. Similar equations hold for irrelevant and correct strategies. Hence:

$$f_1 = w + \tfrac{1}{2}i$$

and, by the same argument:

$$f_j = (\tfrac{1}{2})^j i$$

for j greater than 1. Thus it is shown that the random-sample-of-strategies model with fixed sample size n is a system of recurrent events (this point is obvious and is not proved above) with the same f distribution found in the one-strategy-at-a-time and the all-strategies-at-once models. At this point one can see that the one-strategy-at-a-time model is a special case of the random-sample-of-strategies model with sample size $n = 1$; and similarly, the all-strategies-at-once model is a special case with sample size $n = N$.

One further generalization is possible. Since the final equations of the random-sample-of-strategies model contain no reference to the sample size n, one would get a system of recurrent events with the given f distribution from any mixture of systems which use different sample sizes. One can withdraw the assumption that the sample size is fixed, permitting the possibility that samples are of all different sizes. The only restrictions are that for any n, all samples of size n are equally likely; and there are no empty samples.

The strategy-sampling model differs from Estes' stimulus-sampling model (Estes, 1959) in an important respect. Let the total number of elements (strategies or, in Estes' case, stimulus elements or patterns) be N and the sample size be n. In Estes' formulations, n/N is analogous to the learning rate θ and controls the rate of learning and all the statistical characteristics of the acquisition data. In the present model, learning is independent of the sample size n/N and depends solely on the composition of the basic set H; namely, the proportions of correct, wrong, and irrelevant strategies in H. Thus while the sampling process invoked in this strategy-sampling theory is one of those used by Estes, its place in the theory is entirely different.

In this section it has been shown that three models of the selection of strategies all lead to the same system of recurrent events. Whether or not this is of interest depends largely on whether the resulting description fits the data of cue learning experiments, a question which is considered in the remainder of the paper.

STATISTICAL PROPERTIES OF THE RAW DATA

The discussion above has been restricted to the distribution f_j, the

probability that an error follows the last error by exactly j trials. The f distribution can be estimated directly from raw data, but this is neither a conventional nor a very interesting way of describing the data of cue learning. In this section the data generated by this model are analyzed in several of the ways commonly employed by experimenters. Details are given in Restle (1961).

Learning Curve(s)

Consider three versions of the learning curve. One is the succession of correct and wrong responses by an individual subject. A second is the average learning curve of a group of subjects. The third is a corrected or idealized form of the learning curve computed by adjusting a group learning curve (Vincentizing) or by averaging the data of subjects who are selected after the fact for similarity of over-all performance (Spence, 1956).

According to the present theory, the individual data are composed of a sequence of correct and wrong responses in irregular order, followed by an infinite sequence of correct responses. If $w > 0$, there will be somewhat more errors than correct responses on the resolution trials, and there will be a tendency for errors to follow other errors more often than errors follow correct responses. The probability of an error following an error will be $(w + \frac{1}{2}i)/(w + i)$ whereas the probability of an error following a correct response is $\frac{1}{2}$. An individual subject will produce such below-chance behavior for some block of trials and then abruptly, after some error, will either happen on a correct strategy (in the one-strategy-at-a-time model) or will begin a process of elimination which ends up with all correct strategies. This is an extreme form of the "discontinuous" or insightful learning curve. Unfortunately it is difficult to decide whether any individual subject does or does not exhibit this pattern, so data are usually combined.

The group learning curve is merely the average of a set of (theoretically) discontinuous individual curves. If all subjects have the same parameters c, w, and i, they will nevertheless happen to master the problem at different trials, so that the average learning curve is gradual. Its mathematical form is complex but in general appearance, the group learning from this theory resembles the common "growth" curve (Restle, in 1961).

If one selects, from a larger group of subjects, those who make the same number of total errors, or those who reach criterion at the same time, and averages their learning curves; or if one rescales learning trials as by the Vincent-curve method, the resulting curve will usually be **S** shaped. The writer has investigated this question by generating data which arise directly from the assumptions of the present model, by use of tables of random numbers. When, from a large set of such data, one selects a subgroup of "subjects" who all make the same number of errors, or who reach criterion at about the same trial, and average the performance within such a subgroup, the result is an **S** shaped curve. Exact Vincentizing produces a flat (stationary) learning curve before criterion.

The reason for the **S** shaped curves is not difficult to find; the above methods of selecting or rearranging the data tend to put the (theoretically random) times of solution close together. If the times of solution are put exactly together a step-function should result, but if the times of solution are only grouped close to one

another, the step is blurred and an **S** shaped curve results (Spence, 1956).

Summary Statistics of the Data

In the present theory the actual trial on which learning takes place is not in any way fixed, but depends upon the random outcome of the process of selecting strategies. The result is large intrinsic variance in the acquisition phase, which can most conveniently be described in terms of the total errors made by each subject. If each subject is trained until a long sequence of consecutive correct responses has been obtained, one is reasonably sure that the number of errors made approximates the theoretical total errors.

On the first trial or after any error the subject may either make another error, sometime later, or he may make no more errors. The probability that the subject will make at least one more error, the first one in exactly j trials, is f_j. Thus the total probability of at least one more error is:

$$\sum_{j=1}^{\infty} (f_j) = 1 - c$$

With probability c the subject never makes another error at all. From this it is not difficult to show that the probability of exactly k errors is $(1 - c)^k c$. This is the geometric distribution which has mean:

$$\mathbf{E}(k) = (1 - c)/c$$

and variance:

$$\mathrm{Var}(k) = (1 - c)/c^2.$$

The standard deviation of the distribution of error scores should be nearly equal to its mean, according to the theory, and the distribution should show an extreme positive skewness.

Provided that irrelevant strategies are correct just half the time at random, trials-to-criterion behaves very much like total errors.

Methods of Estimating the Parameters w *and* c

An important step in any mathematical development of a learning theory is the estimation of parameters. Fortunately, in the present model quite simple and efficient estimates are available. A maximum-likelihood estimator of c is given by:

$$\hat{c} = 1/(\bar{T} + 1)$$

where $\bar{T}$ is the mean total error score of a group of subjects. The variance of this estimate, with N subjects in the group, is:

$$\mathrm{Var}(\hat{c}) = c^2(1 - c)/N$$

It is also possible to estimate w by the maximum-likelihood method. As was mentioned earlier, a high frequency of consecutive errors in the presolution phase is an indication of a relatively large proportion of wrong strategies, whereas a chance frequency of consecutive errors is an indication that there are relatively few wrong strategies; provided that the irrelevant strategies are correct and wrong strictly at random. The method is to count "Trial 0," an imaginary trial before training begins, as an error. Then for each subject divide this expanded set of errors into M_0 errors which are followed by correct responses and M_1 errors which are followed by errors. Then computing the means of these statistics one has:

$$\hat{w} = \frac{\bar{M}_1 - \bar{M}_0 + 1}{\bar{M}_0 + \bar{M}_1}$$

as a maximum-likelihood estimator of w (Restle, 1961).

SOME VARIATIONS OF THE BASIC MODEL

The theory stated above is so restrictive in its assumptions that it cannot be applied with success to a great many experimental data. For example, with the theory above, one must expect the standard deviation of error scores to be just slightly less than their mean, and this is not observed with any great regularity. Furthermore, animal subjects often show a strong tendency to remain with one strategy even if it is wrong or irrelevant, for a fairly long string of trials. The model given above says that the subject starts over after each error, and this simply is not a plausible assumption for animal studies.[2]

Perseveration

The model is quite easily generalized to take some account of perseveration of strategies. In place of the assumption that the subject always resamples after an error, one supposes that resampling occurs with some probability r, which may be considered a constant for given experimental conditions and also a constant during the presolution period. The subject chooses some strategy or sample of strategies at Trial 1, whence the probability of making zero errors is c. The probability that any actual error is the last one is only rc, the joint probability that (*a*) the subject resamples and (*b*) having resampled, hits on a correct strategy (in the one-strategy-at-a-time model) or goes into a terminal process of narrowing down (in the all-strategies-at-once or the random-sample-of-strategies models.) For this model with perseveration the distribution of total error scores is:

$$P(0) = c$$
$$P(T) = (1 - c)(1 - rc)^{T-1}rc.$$

where $T > 0$. The value of this model in fitting animal data is illustrated by analyzing a set of error scores reported in detail by Harlow, Harlow, Rueping, and Mason (1960), (See Restle, 1960). Harlow's distributions show a relatively high frequency of solutions with zero errors, along with a spread to very large error scores, in original learning by baby monkeys. The result is that the standard deviations of the distributions are somewhat larger than the means. Harlow et al. remarked on the high degree of perseveration shown by some monkeys. Even more striking is Warren's investigation (Warren, 1959b) of the discrimination learning of cats, which show strong perseverative tendencies and which give distributions of error scores with standard deviations much larger than the corresponding means.

It is not at all sure that perseverative effects can be reduced to the single parameter r, and the writer's investigations are entirely insufficient to support the assumption. Within the model there are several symptoms of perseveration; a high frequency of zero error scores relative to the remainder of the distribution, higher standard deviation than mean errors, and consecutive runs of responses which follow a wrong or irrelevant strategy, such as a position habit, object preferance, etc. One can imagine that perseverative tendencies would be strong whenever the incentive for learning is not entirely effective, or when the subjects (perhaps from partial reinforcement) have developed an expectation that rewards cannot always be attained. Something of this last effect, in college students, is

[2] The writer is indebted to Marvin Levine for pointing out the importance of this factor in animal discrimination learning.

suggested by the results of Morin (1955).

Complexity of Correct Strategies

In the model above it is tacitly assumed that the problem can be solved on the basis of a single correct strategy. This seems a reasonable assumption in the case of simultaneous discrimination. However, when two situations are presented on successive trials it is a question whether complete solution involves learning one strategy (turn right to white and left to black) or two strategies, one for each situation. If two strategies are learned independently and with the same parameters, then the distribution of errors is the distribution of the sum of two random variables (total errors in each situation) each of which has a geometric distribution. This results in a negative binomial distribution (see Restle, 1961) with mean:

$$\mathbf{E}(T) = 2(1 - c)/c$$

and variance:

$$\mathrm{Var}(T) = 2(1 - c)/c^2.$$

These statements follow at once from the fact that the mean of the sum of two independent variables is the sum of the means, and the variance of the sum is the sum of the component variances. With small c the standard deviation will be about $1/\sqrt{2}$ or about 0.7 of the mean.

An interesting fact can be brought forward in favor of this argument, though the writer has not been able to make an exhaustive test. In a case of successive size discrimination by college students (Restle, 1955) the writer found ratios of standard deviation/mean of 1.1 and 0.9 for two problems—both very close to unity, hence in agreement with the original model and with the notion that a single strategy is sufficient to solve the problem. In a comparable study using rats in a successive black-white discrimination, Amsel (1952) obtained data on three groups in which the ratio of standard deviation to mean was 0.65, 0.68, and 0.72. The rat results are in excellent agreement with the hypothesis that the rats had to use two separate strategies to solve the problem. One might guess that rats do not integrate the two habits, right turn to white and left turn to black, into a single cognitive structure, whereas humans might make just such an integration. The present analysis may be useful in throwing further light on the nature of successive discriminations for various species of subjects.

ADDITIVITY OF CUES

In several papers regarding another theory the writer has discussed the additivity of cues (Restle, 1955, 1957, 1958, 1959a; see also Bourne & Restle, 1959, and Trabasso, 1960). In simple terms the experiment involves three groups; one learns a problem based on a set A of cues, the second learns a problem based on the set B of cues, and the third learns a problem which can be solved using either A or B cues disjunctively. If the sets A and B are separate then the third set $A \cup B$ should have measure $m(A \cup B) = m(\mathrm{A}) + m(\mathrm{B})$, (see Trabasso, 1960). Experimental results have been reconciled with an S-R theory involving adaptation of irrelevant cues (Restle, 1955).

Generally speaking, the calculations from the present model are in good numerical agreement with those from the "adaptation" model. Reanalysis of three of the most satisfactory sets of data used before, and one set not previously used, are reported

here. The other data previously discussed are so fragmentary that analysis will hardly be fruitful; and Trabasso (1960) used such difficult problems that many of his subjects failed to learn at all on the more difficult problems, making analysis by the present model unsatisfactory. In general, the estimates reported below are not the maximum-likelihood estimates because the subjects were not run to a strong criterion. However, since learning was nearly complete, the approximate estimates are adequate.

Scharlock (1955; see also Restle, 1957) ran a place-versus-response experiment with rats in which the relative weight of correct place (extra-maze) and correct response (intra-maze) strategies can be estimated. He also ran one group with both place and response strategies correct, and a group with response strategies correct and no place cues present. In the calculations below it is assumed that the number or wrong strategies of a given type (place or response) equals the number of correct strategies. This is a symmetry assumption which, while not strictly appropriate for Scharlock's experiment, is needed to permit prediction, since separate estimates of wrong strategies cannot be made on the available data.

Rats learned to go to the same place (using extra-maze cues), with response cues irrelevant, making an average of 9.7 errors. Using the estimate $\hat{c} = 1/(1 + \bar{T})$, one estimates that the correct place strategies make up $1/10.7 = .093$ of the total set of strategies. Other rats learned to make a constant response to different places, averaging 6.7 errors. We estimate that correct response strategies constitute $1/7.7 = .130$ of the total set.

The simplest interpretation of the experiment is that when both place and response strategies would work, the proportion of correct strategies would be the sum of the proportions of place and response strategies, since the same total set of strategies is available. One predicts that with place-plus-response learning, $c = .093 + .130 = .223$. We have the formula that the expected errors to solution is $(1 - c)/c$. Hence the expected mean errors is $.777/.223 = 3.4$; the obtained mean was 4.0, which is adequately close. The discrepancy can partly be explained by the fact that the place-learning and response-learning groups were not run to a high criterion and probably would have made more errors if tested longer. The fast-learning place-plus-response group would likely not have made more errors, since their performance was excellent at the end of the training given.

Another group learned a fixed response in the absence of any good place cues. We assume that this group simply had neither correct nor incorrect place strategies. Its predicted proportion of correct strategies is $.130/1 - 2(.093) = .130/.814 = .160$. The expected errors is $.840/.160 = 5.25$; the observed mean was 5.0. Both predictions are close to the obtained results, well within sampling deviations.

Warren (1959a) has reported data on monkeys in an experiment analogous to Scharlock's study of rats. Warren's monkeys had to learn position habits in some problems and object-discriminations in other problems. Warren also used a response-plus-object problem (object discrimination with the objects left in the same place each trial) and a pure response problem (e.g., choose the left one of two indistinguishable objects.) Predictions followed the same formulas as above, and were extremely accurate: the proportions of object and

response strategies were estimated from behavior on object (position-varied) and position (object-varied) problems. When these values were added to predict object-plus-position, the prediction was a mean of 0.67 errors, whereas Warren observed 0.63. For pure response learning, the model predicted 3.25 mean errors and the observed value was 3.04. These agreements between theory and data are well within the range of probable sampling error.

Similar results were obtained by analyzing Warren's experiment on the additivity of color, form, and size cues (Warren, 1953). The analysis is substantially the same as that given in a previous paper (Restle, 1958). Learning data are collected on problems involving only color, only form, or only size differences between the objects. These data are used to calculate the proportional weights of the three sources of strategies, and the resulting values are recombined to compute predicted learning rates for problems involving two or more dimensions; for example, discrimination of a red triangle from a green circle involves color plus form cues. In each calculation account is taken of the greater total number of strategies involved in problems with added cues, and it is assumed that whenever a set of correct strategies is introduced by adding cues, an equally large set of wrong strategies also enter the situation.

Calculations on Warren's (1953) data afforded four predictions of total errors to solution. The four predictions were wrong by -17, 0, 5 and 10 percent respectively, and all of the errors can reasonably be attributed to sampling variations.

The hypothesis of additivity of cues has also been applied to human learning (Restle, 1959a) in an experiment which required subjects to learn differential verbal responses to consonant syllables. The procedure was simple concept formation with individual letters of the syllables used as cues, and the same sort of additivity of cues as above. The adaptation model gave quite accurate predictions and the present model is, if anything, slightly more accurate; it predicts (given data on the two cues separately) that the added-cue group should make an average of 5.26 errors, whereas the observed mean was 5.25. Considering the variability of the data, the extreme closeness is coincidental.

COMPARISON WITH OTHER THEORIES

The three models of the selection of strategies are conceptually similar to the writer's adaptation theory of discrimination learning (Restle, 1955). The "strategies" of the present model resemble the "cues" of the earlier theory. The strategy-selection model has, as a theorem, that the rate of learning depends upon the proportion of correct strategies. A similar idea is expressed in the adaptation model in which it was assumed that the rate of learning (θ) would depend upon the proportion of relevant cues.

Several of the serious faults of the adaptation theory are corrected in the theory of strategy-selection. First, in the adaptation theory the subject was supposed to begin conditioning relevant cues and adapting irrelevant cues, with a rate θ equal to the proportion of relevant cues, right from the first trial of training. Of course, at that first trial the subject has no possible way of knowing which cues will turn out to be relevant—it would be possible for the experimenter to change his mind after the first trial is complete. Hence, the adaptation theory could apply only to a prescient subject. The selection-of-strategies

theory does not have the subject treat different strategies differently except on the basis of trials already completed, hence avoids the absurdity. Second, the idea that the structure of the problem (proportion of relevant cues or correct strategies) controls the rate of learning was only a simplifying assumption, with no justification, in the adaptation model; but is an inescapable theorem of strategy-selection. Third, the adaptation theory yielded a determinate learning curve in the sense that $p(n)$ was exactly specified for a given θ and $p(1)$. The variance of typical data was in large part left unexplained by the adaptation model and had to be attributed to individual differences, even though it is notoriously difficult to find any strong predictor of such learning. In the strategy-selection model learning itself is a random event and the model generates variability comparable with that obtained in the data.

Despite these important differences, the present theory is close enough to the adaptation theory to make it possible to carry over many of the theoretical insights, though in modified form. The one type of prediction discussed in this paper was additivity of cues, but other ideas such as the proposed basis of learning sets (Restle, 1958) and various quantitative relationships in concept identification (Bourne & Restle, 1959) can be recast in the mold of the selection of strategies. The many similarities and few differences in predictions must be studied in detail and cannot be discussed here, except to mention that most of the results of the Bourne and Restle paper can be reproduced using the strategy-selection model.

In more general terms, it may be remarked that the strategy-selection model is similar to the ideas of Lashley (1928) and Krechevsky (1932) in general intent. In comparison with stimulus-sampling theoreies (Restle, 1959b) the strategy-selection model is like theories of the observing response (Atkinson, 1959a, 1959b; Wyckoff, 1952), except that in the strategy-selection model there is no conditioning, only observing (selecting). The mathematical structure of the model is very close to that of Bower's (1960) one-element association model.

SUMMARY

A stochastic model for the solution of cue learning problems by the selection of strategies was stated and developed. Errors were shown to constitute a system of uncertain recurrent events in Feller's sense. Three models, one-strategy-at-a-time, all-strategies-at-once, and a-random-sample-of-strategies, were formulated and shown to yield the same system of recurrent events, to be identical in terms of data. This is the independence-of-sample-size theorem. The basic distribution of the recurrent-events system was used to generate a description of the data, and also a method for estimating the proportions of correct and wrong strategies in a problem. Variations of the model, which take account of perseveration and complex strategies, were indicated. Experimental evidence, mainly on the additivity of cues, was discussed.

REFERENCES

AMSEL, A. Rate of learning a visual brightness discrimination as a function of discriminanda durations. *J. comp. physiol. Psychol.*, 1952, **45**, 341–346.

ATKINSON, R. C. The observing response in discrimination learning. Technical Report No. 4, 1959, University of California, Los Angeles, Contract Nonr 233(58). (a)

ATKINSON, R. C. A theory of stimulus discrimination learning. Technical Report No. 1, 1959, University of California, Los Angeles, Contract Nonr 233(58). (b)

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BOWER, G. H. Properties of the one-element model as applied to paired associate learning. Technical Report No. 31, 1960, Stanford University, Contract Nonr 225(17).

ESTES, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959.

FELLER, W. *An introduction to probability theory and its applications*. (1st ed.) New York: Wiley, 1950.

HARLOW, H. F., HARLOW, M. K., RUEPING, R. R., & MASON, W. A. Performance of infant Rhesus monkeys on discrimination learning, delayed response, and discrimination learning set. *J. comp. physiol. Psychol.*, 1960, **53**, 113–121.

KRECHEVSKY, I. "Hypotheses" in rats. *Psychol. Rev.*, 1932, **39**, 516–532.

LASHLEY, K. S. *Brain mechanisms and behavior*. Chicago: Univer. Chicago Press, 1928.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 343–351.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Discrimination of cues in mazes: A resolution of the "place-vs.-response" question. *Psychol. Rev.*, 1957, **64**, 217–228.

RESTLE, F. Toward a quantitative description of learning set data. *Psychol. Rev.*, 1958, **65**, 77–91.

RESTLE, F. Additivity of cues and transfer in discrimination of consonant clusters. *J. exp. Psychol.*, 1959, **57**, 9–14. (a)

RESTLE, F. A survey and classification of learning models. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory*. Stanford: Stanford Univer. Press, 1959. (b)

RESTLE, F. A note on the "hypothesis" theory of discrimination learning. *Psychol. Rep.*, 1960, **7**, 194.

RESTLE, F. Statistical methods for a theory of cue learning. *Psychometrika*, 1961, **26**, 291–306.

SCHARLOCK, D. P. The role of extramaze cues in place and response learning. *J. exp. Psychol.*, 1955, **50**, 249–254.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

TRABASSO, T. R. Additivity of cues in discrimination learning of letter patterns. *J. exp. Psychol.*, 1960, **60**, 83–88.

WARREN, J. M. Additivity of cues in visual pattern discriminations by monkeys. *J. comp. physiol. Psychol.*, 1953, **46**, 484–486.

WARREN, J. M. Solution of object and positional discriminations by monkeys. *J. comp. physiol. Psychol.*, 1959, **52**, 92–93. (a)

WARREN, J. M. Stimulus perseveration in discrimination learning by cats. *J. comp. physiol. Psychol.*, 1959, **52**, 99–101. (b)

WYCKOFF, L. B. The role of observing responses in discrimination learning. Part 1. *Psychol. Rev.*, 1952, **59**, 431–442.

(Received November 14, 1960)

Psychological Review
1962, Vol. 69, No. 4, 344–354

SUBJECTIVE ORGANIZATION IN FREE RECALL OF "UNRELATED" WORDS [1]

ENDEL TULVING

University of Toronto

This paper is concerned with organization as a dependent variable in free recall verbal learning. A method will be described for measuring the extent to which subjects' recall of verbal items presented in different orders on successive trials is structured sequentially. The method assumes no knowledge on the experimenter's part as to the sources of organization, and it is thus applicable to free recall learning of any list of verbal items. The effect of repetition on organization, and the relation of organization to amount of recall will be briefly examined in a simple experiment. The method and the data are discussed in terms of the problem of the role of repetition in free recall verbal learning.

Verbal Learning and Organization

The fact that learning occurs under conditions of practice, often apparently through mere repetition of the material, is empirically well known, but theoretically not yet fully understood. Indeed, the question of why repetition leads to better recall has not been raised too often explicitly, although it has been implicit in much experimental and theoretical work.

The apparent lack of explicit interest in this problem seems to be related to the restricted aspects of subjects' behavior which are studied in verbal learning experiments. In a large majority of studies the dependent variables have been based on the operation of counting single unordered responses, assigned to categories such as "correct," "intralist intrusion," "remote backward association," and the like. Given but a single basic response variable of this kind, there is relatively little that can be done with the question of why repetition is effective, other than studying the effect of different independent variables on the rate of learning. One can be justifiably sceptical, however, of our chances of ever understanding the basic process of verbal learning by simply counting "correct" and various classes of "incorrect" responses under a variety of experimental conditions. Bolles' (1959) recent suggestion that, "If we are to understand human learning, we must determine what our *S*s are really doing" (p. 580), can be regarded as an example of growing realization that a critical reappraisal of response variables in verbal learning is overdue.

An important contribution to our understanding of the basic acquisition process might well be provided by response variables based on ordered classifications of data (Miller & Frick, 1949). Such variables are particularly interesting in the light of G. A. Miller's (1956a, 1956b, 1956c), conception of verbal learning as a direct consequence of the process of recoding or organization. Rehearsal or repetition, according to Miller, 1956b), has "the very important effect of organizing many separate items into a single unit" (p. 43). Repetition does not change the basic storage

[1] This research has been supported by the National Research Council of Canada under Grant Number APA-39.

capacity of memory. Rather, organizing processes accompanying repetition lead to an apparent increase in this capacity by increasing the information load of individual units.

Miller's unitization hypothesis, if found tenable, has important implications for research strategy and theory in verbal learning. One of the difficulties in evaluating the hypothesis lies in the lack of appropriate measures of organization. Organization as a response variable has been investigated by many psychologists (e.g., Bousfield, 1953; Cofer, 1959; Jenkins & Russell, 1952; Rothkopf & Coke, 1961), but its measurement has always depended on the experimenter's knowledge of sources of organization present in the stimulus list. It is well known, however, that subjects are quite capable of memorizing materials which are not organized in any obvious manner, or for which the sources of organization cannot be readily specified. Thus, if the concept of organization is to have more general applicability, it seems necessary to develop methods that permit observation and quantification of organization as a response measure independent of the characteristics of the stimulus material. This is what the present paper attempts to do.

Behavioral manifestations of the hypothesized organizing process can best be studied under conditions where the order in which the subject recalls items is free to vary. Items which are organized into a single unit would then be expected to occur in close temporal contiguity in subject's recall. Repeated occurrences of such sequential patterns would indicate the existence of the organizing process, and the extent to which this occurs would provide an estimate of the degree of organization. Thus a quantitative analysis of sequential dependencies among items in subject's free recall on successive trials would yield measures of a response variable closely related to the hypothetical organizing process, independently of quantity of recall.

The method proposed in this paper is derived from information theory (Shannon & Weaver, 1949), and is quite similar to Miller and Frick's (1949) "index of behavioral stereotypy." It provides for a measure of sequential redundancy in repeated ordered samples of a set of items. As the experimental paradigm used here involves presentation of material completely free from any sequential redundancy, such redundancy or organization in subject's recall cannot be attributed to the input material. Rather, it is the subject who imposes a certain degree of organization on the material. For this reason such organization is called subjective organization (SO). In information theory terms, SO is "noise" produced by the channel, that is the subject; it is information in the output not found in the input (Miller, 1953).

MEASUREMENT OF SUBJECTIVE ORGANIZATION

Suppose we select L words, any words, and construct L different orders of these words such that each word appears in each of the L serial positions once and is preceded and followed by each other word just once. This can be done easily if $L + 1$ is a prime number. We then represent each of these L ordered lists to the subject on L separate trials. Words are shown one at a time on the memory drum. After each trial the subject is asked to record all the words that he remembers from the list, in any order he wishes. After all L ordered lists have thus been presented we have L recall records from the subject, one

TABLE 1

A SAMPLE RECALL MATRIX FOR AN INDIVIDUAL SUBJECT

nth word	$(n+1)$th word																	
	y_0	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	n_i
x_0		1	1	2	2	1	3				2	1	1			1	1	16
x_1							1		1		9	1			2			14
x_2	1	2			1	2						2			2	2		12
x_3	2	1			1	2	1	2				2	1			1		13
x_4	1	2		1			1	5				3		1	1			15
x_5	1						2		1	8				1		2		15
x_6	1	1		1					1			1		11				16
x_7	1		1		2								1		7			12
x_8	2	2				3					2		1			4		14
x_9		1			1	1			10						1	1		15
x_{10}		1	1		1	1	3			1		1	2			1	1	13
x_{11}	1				6	1	2										4	14
x_{12}	2	1						2				1		1	1	1	4	13
x_{13}		2		9		1	2			1		1						16
x_{14}	1		9					1					2	1				14
x_{15}	2				1	3	1		1	5				1				14
x_{16}	1							2				1	5			1		10
n_j	16	14	12	13	15	15	16	12	14	15	13	14	13	16	14	14	10	$N = 236$

Note.—Recall is pooled over 16 trials.

for each trial. We then tabulate the frequency of all pairs of adjacent recall responses in a matrix consisting of $L + 1$ rows and $L + 1$ columns, and compute the measure of SO from the data in the matrix.

Table 1 shows a sample recall (output) matrix for a case where $L = 16$ and where the subjects recall responses are pooled for all 16 trials. Symbols $x_1, x_2, \ldots x_{16}$ identifying the rows refer to the 1st, 2nd, . . . 16th word in the nth position in the subject's recall list; symbols $y_1, y_2 \ldots y_{16}$ identifying the columns stand for the same words in the $(n + 1)$th position in the subject's recall list. Thus x and y represent successive positions in the subject's recall, and subscripts identify the contents of these positions, i.e., words. Symbol x_0 stands for the blank position (no word) immediately preceding the first word in the subject's recall list, and y_0 refers to the blank position immediately following the last word in the same list. The entries in the cells of the matrix show the frequency with which word i was followed by word j in the subject's recall on all 16 trials; their numerical values are symbolized as n_{ij}. Entries in the row labelled x_0 show the frequencies with which each word j stood at the beginning of the subject's recall lists on 16 trials, and similarly entries in the column y_0 show frequencies with which each word i was in the last position of the recall lists on these trials. Zeros have been omitted in the matrix. For instance, $n_{06} = 3$, which means that Word Number 6 was in the first position (followed no other word) in the subject's recall on three trials; $n_{13} = 0$, indicating that Word Number 3 never followed Word Number 1 during the 16 trials, and so forth. The numerical values of mar-

ginal totals of rows, n_i, show how many times each word i appeared in the subject's recall on all 16 trials. Marginal totals for corresponding rows and columns are of course identical. It is also to be noted that there are no entries in the cells along the main diagonal, from upper left to lower right, since normally no word follows itself in recall.

Given a recall matrix such as the one in Table 1, a number of different measures of second-order sequential organization could be used. The measure of sequential organization adopted in the present paper is essentially a measure of redundancy in a sequence of events when the probabilities of successive pairs of events are known or can be estimated:

$$C_x(y) = 1 - \frac{H_x(y)}{\max H_x(y)} \qquad [1]$$

$H_x(y)$ here refers to the amount of information (uncertainty) in the event y when the preceding event x is known, and $max\ H_x(y)$ stands for the maximum value of such uncertainty. This is Miller and Frick's (1949) second-order index of behavioral stereotypy. The formula for computing second-order SO involves a minor modification of the above formula,[2] simplifying calculations in addition to being somewhat more appropriate to the free recall learning situation. This formula is as follows:

$$SO = \frac{\sum_{i,j} n_{ij} \log n_{ij} - \min \sum_{i,j} n_{ij} \log n_{ij}}{\sum_{i} n_i \log n_i - \min \sum_{i,j} n_{ij} \log n_{ij}} \qquad [2]$$

where n_{ij} represents the numerical value of the cell in the ith row and jth column, n_i represents the marginal total of the ith row, and $\min \sum_{i,j} n_{ij} \log n_{ij}$ represents the minimum value that $\sum_{i,j} n_{ij} \log n_{ij}$ can assume.

In cases where no marginal total n_i exceeds the length of the list L, $\min \sum_{i,j} n_{ij} \log n_{ij}$ is equal to 0, and Formula 2 reduces to the following relatively simple expression:

$$SO = \frac{\sum_{i,j} n_{ij} \log n_{ij}}{\sum_{i} n_i \log n_i} \qquad [3]$$

In this formula $\sum_{i} n_i \log n_i$ can be regarded as a measure of maximum organization, $\sum_{i,j} n_{ij} \log n_{ij}$ represents the actual organization, and SO is simply a measure of actual organization relative to the maximum. Thus SO can assume all values between zero and unity, the former expressing the complete absence of second-order sequential organization, as in the case of the ordered stimulus lists presented to the subject, and the latter the maximum degree of such organization, as in the case of an imaginary subject who recalls all words on all trials exactly in the same order. SO for the data in Table 1, incidentally, is .397.

So far we have considered only second-order sequential organization,

[2] *Max* $H_x(y)$ in Miller and Frick's index of behavioral stereotypy is defined as $max\ H(x,y) - max\ H(x)$. Thus it is based on the assumption that all responses are equally frequent. For the present problem, however, it may be somewhat more meaningful to compute $max'\ H_x(y)$ on the basis of the *given* frequency of individual responses, i.e., as determined by subject's recall:

$$max'\ H_x(y) = max\ H\ (x,y) - H\ (x)$$

If $max'\ H_x(y)$ is substituted for $max\ H_x(y)$ in Formula 1, that formula becomes mathematically equivalent with Formula 2, the measure of subjective organization as used in this paper.

based on pairs of successive responses. The method can be easily extended to higher-order dependencies, even though the computational labors involved would become quite prohibitive. However, estimates of higher-order organization can be obtained more easily. For instance, instead of setting up a three dimensional matrix for calculating third-order SO, we can tabulate data in a two dimensional matrix with nth items as rows and $(n + 2)$th items as columns. In this case we also need two rows for blank spaces preceding the first word (x_0 and x_{0-1}), and two columns for blank spaces following the last word in each recall list (y_0 and y_{0+1}). This procedure amounts to collapsing the three dimensional matrix along the $(n+1)$th dimension. That is, we are computing an SO measure based on pairs of responses *separated* by one response. Such an SO is labeled SO (Lag 1), and it does provide an estimate of the third-order SO. In a similar manner, still higher order dependencies can be evaluated. SO (Lag 2) is an estimate of the fourth-order sequential organization, SO (Lag 3) is an estimate of the fifth-order SO, and so forth.

SO can be calculated for any block of trials, usually successive trials. Thus it is sometimes useful to indicate the size of the block on which the measure is based. For instance, SO (Blks = 3) means SO calculated from blocks of three successive trials. SO for a single trial is not meaningful, as subjective organization is defined in terms of the subject's tendency to recall words in the same order on successive trials; in fact SO for a single trial, when calculated according to the formulas given, turns out to be 0, provided that the subject has not repeated any pairs of words.

DEMONSTRATION EXPERIMENT: SUBJECTIVE ORGANIZATION AND PERFORMANCE

The following simple experiment will serve as an illustration of how the method works, and the data from the experiment will provide tentative answers to three questions which can be raised in the context of the discussion of the concept of organization in the first section. The questions are as follows: First, do the subjects actually organize "unrelated" words when instructions are simply to recall as many words as they can? Second, does this subjective organization , if found in the data, increase with repetition of the material? Third, is there a systematic relation between subjective organization and more traditional measures of learning, such as the number of words correctly recalled?

Procedure

A list of 16 English words was adopted from another verbal learning experiment (List III in Tulving & Thornton, 1959). All words were disyllabic nouns, consisting of five, six, or seven letters. The words, here listed alphabetically, were as follows: ACCENT, BARRACK, DRUMLIN, FINDING, GARDEN, HOYDEN, ISSUE, JUNGLE, LAGOON, MAXIM, OFFICE, POMADE, QUILLET, TREASON, VALLEY, WALKER. Sixteen different sequences of these words were constructed such that, considering any block or all 16 orders of the stimulus list, there was no second-order or higher-order redundancy in the lists.

Sixteen female undergradute students enrolled in the introductory psychology courses at the University of Toronto served as subjects. Their median age was 19 years. None of the subjects had previously participated in any verbal learning experiments.

Each subject was tested individually. In the instructions the subject was told that her task was to learn a list of 16 words, two-syllable English nouns, that would be presented to her on 16 separate trials. At the end of each trial she was to write down as many words from the list as she remembered. Then she would be shown the same words again, in a

different order, and at the end of the trial she would again write down all the words that she was able to recall. The subject was also told that the order in which she recalled the words did not matter, her task was simply to recall as many of the words on each trial as she could.

The 16 different orders of the stimulus list were typed on white paper in lower case letters and presented to the subject on a memory drum. The rate of exposure was one word per second. At the end of the trial the subject was given 90 seconds to record her recall on sheets of paper lined with 16 consecutively numbered lines.

The order in which the 16 ordered lists were presented to the subjects was systematically counterbalanced among subjects. The method used for assigning the 16 orders of the stimulus list to the 16 subjects on the 16 trials was the same as that used in assigning the 16 words to 16 serial positions in 16 orders of the stimulus list.

Results

Two kinds of data are of interest in this experiment, performance (P) defined in terms of frequency of correct recall, and subjective organization (SO). Extralist intrusions and misspelled words from the list were ignored in computing both P and SO scores.[3] Repetitions of list words within a given trial were included in the recall matrix and thus entered the SO score, but not the P score. The mean number of extralist intrusions, together with misspelled words, was 0.13 per trial, the mean number of repetitions of list words was 0.09 per trial.

Second-order SO scores were calculated for successive blocks of two trials (SO, Blks = 2), for running blocks of three trials (SO, Blks = 3), and for the total block of 16 trials (SO, Blks = 16). SO (Blks = 2) scores were also computed for 16 statistical subjects.[4]

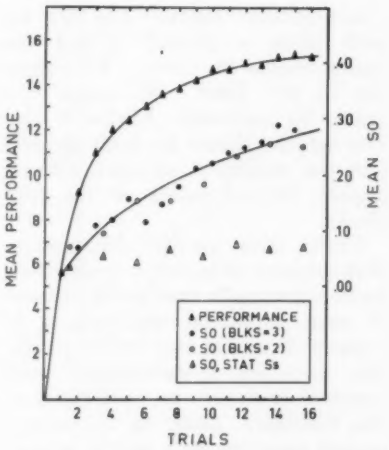

FIG. 1. Mean performance (upper curve) and mean SO (lower curve) as a function of trials. (Values of performance are to be read from the left ordinate, SO from the right ordinate. Mean SO for blocks of two trials from 16 statistical subjects are shown for comparison, open triangles. For further explanation see the text.)

The effect of repetition on both P and SO is summarized graphically in Figure 1. The values of P are to be read from the left, the values of SO from the right hand ordinate. The abscissa represents trials. The upper curve is the typical learning curve, showing performance as a function of trials. The lower curve, fitted by inspection to data points representing SO (Blks = 3), could be called an

[3] Such extralist intrusions are of interest in their own right and may provide valuable information about the organizing process. In order to handle them in the present situation, the recall matrix can be simply extended. However, such intrusions can probably be better studied in different experimental situations (e.g., Deese, 1959).

[4] IBM Model 650 Electronic Data Processing Machine was used in these calculations. Thanks are due to Albert S. Bregman for constructing the program. Data from statistical subjects, whose "performance" was matched with that of individual experimental subjects, were collected as an attempt to estimate the amount of organization occurring by chance, since the sampling distribution of SO, under the null hypothesis of no organization, is not known.

"organization" curve. The SO for each block is plotted against the middle trial of the block. Thus there are 14 SO (Blks = 3) scores. SO scores for successive blocks of size two are also shown for both the real and the statistical subjects, plotted against the midpoints of the trial blocks.

These data clearly demonstrate that subjects do in fact organize their recall sequentially even in the absence of such sequential organization in stimulus lists, and that this organization increases systematically with repeated exposures to, and recall of, the material. Thus, in an experimental situation such as the present one, repetition has two parallel effects: increasing frequency of recall is accompanied by an increasingly tighter sequential organization.

The slopes of the two curves depicted in Figure 1 seem to be different. When mean P scores are plotted against mean SO (Blks = 3) for the 14 trials, the curvilinearity of the relation is obvious. However, the relation becomes quite linear when a logarithmic transformation is applied to the mean SO scores, suggesting that, as a first approximation, increase in performance is proportional to increase in log SO. The product-moment correlation between mean P and log mean SO (Blks = 3) for Trials 2 to 15 was found to be +.96.

Estimates of higher-order SO were obtained in the form of SO scores for Lags 1, 2, and 3, for the total block of 16 trials. The mean SO (Blks = 16) scores for these lags and their standard deviations are shown in Table 2, together with comparable data from the statistical subjects for Lags 0 and 3. Even though mean SO scores seem to decrease with increasing lags, it is interesting to note that SO (Lag 3, Blks = 16) scores are higher than the same scores from the statistical subjects. The median test yielded a chi-square of 15.12, which is highly significant.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF SO (Blks = 16) FOR VARIOUS LAGS

Group	Lag 0	Lag 1	Lag 2	Lag 3
Experimental subjects				
Mean	.292	.224	.223	.216
Standard deviation	.054	.045	.034	.035
Statistical subjects				
Mean	.180			.169
Standard deviation	.017			.012

Note.—N = 16 for each group.

TABLE 3

INTERCORRELATIONS AMONG SO (Blks = 16) MEASURES FOR LAGS 0, 1, 2, AND 3; AND CORRELATIONS BETWEEN THESE SO SCORES AND MEAN P SCORES ON TRIALS 1 TO 8, 9, TO 16, AND 1 TO 16 FOR INDIVIDUAL SUBJECTS

Measures	SO			
	Lag 0	Lag 1	Lag 2	Lag 3
SO (Lag 1)	+.86			
SO (Lag 2)	+.62	+.85		
SO (Lag 3)	+.58	+.74	+.84	
P (Trials 1 to 8)	+.45	+.32	+.24	+.01
P (Trials 9 to 16)	+.78	+.61	+.34	+.19
P (Trials 1 to 16)	+.63	+.47	+.30	+.08

Note.—N = 16.

The intercorrelations among SO (Blks = 16) scores for Lags 0, 1, 2, and 3, as well as correlations between these SO scores and several average P measures are shown in Table 3. These correlations are based on individual data from 16 subjects.

Certain orderly relations appear in Table 3. First, the correlations between any two adjacent SO measures on the lag dimension are reasonably high, all three listed in the table being approximately +.85. Second, with increasing distance on the lag dimension between any two SO measures the correlations decrease. Thus, for

example, SO (Lag 0) has a correlation of +.85 with that for Lag 1, +.62 with Lag 2, and +.58 with Lag 3. Third, correlations between P and SO decrease with increase in the lag of the SO measure. The most important conclusion to be drawn from these data is that, within the limits of the present method, the second-order SO is about as useful a measure of sequential organization as that based on any combination of different orders, and more useful than any other single higher-order SO measure.

Positive correlations between SO and P scores in Table 3, where correlation is done on data from 16 subjects, support the observation of the same relation over trials. It is also interesting to note that on the basis of a single estimate of organization for each subject, that based on the total block of 16 trials, SO accounts for a larger amount of variance in the P variable for Trials 9 to 16 than for Trials 1 to 8.

All the above findings relate to organization as an intrasubject phenomenon. Commonality of organization between subjects, however, can also be investigated by using the same general method. That there might be such commonality became apparent in the inspection of recall data from different subjects. Recall matrices from individual subjects showed that quite often the patterning of response sequences was similar for many subjects. This phenomenon was investigated in two ways.

First, the recall data from all 16 subjects were pooled for all 16 trials and entered into an "intersubject recall matrix." If all subjects organized their recall differently, the distribution of cell entries in the group recall matrix would not be significantly different from the distribution under conditions where only chance factors

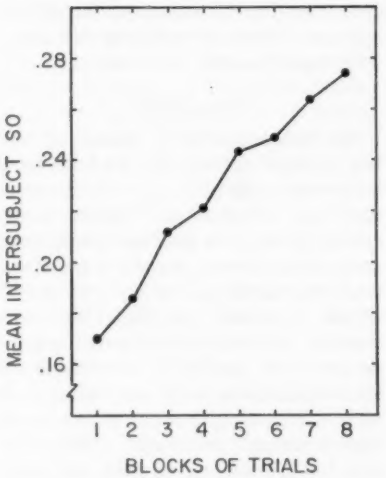

FIG. 2. Mean intersubject SO scores for 16 subjects on eight successive blocks of two trials.

are operating. To test the significance of the apparent deviation of the data in this matrix from chance distribution, the information function $T(x,y)$ was computed. It was found to be 0.352 bits which is significant at better than .01 level, using the method suggested by Miller (1955). This finding thus confirms the casual observation that there is a certain degree of commonality in the subject's recall.

The second method used to explore intersubject organization was as follows. For a single trial data were pooled in a recall matrix from all 16 subjects. For each such intersubject recall matrix SO was computed. This procedure results in 16 intersubject SO measures, one for each trial. These data are shown in Figure 2. Intersubject SO measures have been averaged for blocks of two trials, in order to smooth the curve. In view of the orderly increase in the intersubject SO measure with repetition, it would seem safe to conclude that the

commonality of organized recall sequences between subjects increases with repetition of the material.

DISCUSSION

All three questions posed at the beginning of the experiment have been answered in the affirmative. It seems that the subjects do impose a sequential structure on their recall, that this subjective organization increases with repeated exposures and recall of the material, and that there is a positive correlation between organization and performance. None of these findings is very surprising, and all are in good agreement with related experimental evidence. Bousfield and his associates, as well as other investigators, have amply demonstrated that there is a strong tendency in subjects to recall randomly presented material in sequences of related words or clusters (Bousfield, 1953; Cohen & Bousfield, 1956; Jenkins & Russell, 1952). It has also been shown that repeated presentations of the stimulus word list increases this clustering tendency (Bousfield & Cohen, 1953), and that there is a direct relation between the degree of clustering and the amount of recall (Bousfield & Cohen, 1955; Bousfield, Cohen, & Whitmarsh, 1958; Jenkins, Mink, & Russell, 1958; Sakoda, 1956). The findings of the present experiment extend the domain of these phenomena of organization from one trial recall and experimentally organized materials, as used in all the above experiments, to a learning situation, with several successive trials, and experimentally unorganized materials.

The finding that the organizing effects, well demonstrated in experiments on clustering, can be experimentally assessed even in case of unrelated words, is quite encouraging. Perhaps paradoxically this suggests that a list of completely *unrelated* words is probably as fictional as is a truly nonsensical nonsense syllable. It is for this reason that reference has been made throughout this paper to "unrelated" words, referring only to the fact that words are experimentally unselected as to their meaning.

The present method of quantifying sequential organization in free recall learning constitutes only a rather coarse net for capturing the process of organization in all its diversity. For one thing, the SO measure is most sensitive to rigid sequences of responses, or chunks (Miller, 1956c). It is quite likely, however, that a great deal of organization occurs in the form of clusters (Bousfield, 1953), in which items have no fixed order. This would tend not only to attenuate measures of SO, but also, through increased variability of SO, depress correlations with the performance variable.

It would be interesting, in view of Miller's unitization hypothesis, to look at organization in terms of its units. In many cases subjects undoubtedly impose a hierarchical organization on recall, or use a Plan (Miller, Galanter, & Pribram, 1960). In other cases, particularly when the list to be learned is short and some coding device such as the alphabet can be employed, nonhierarchical organization may also occur. The present method does not lend itself to an identification of various strategies adopted by subjects, even though casual inspection suggests many diverse sources of organization: associative grouping, conceptual categories, assonance, grouping in terms of familiarity of items, and so forth.

The method provides only an overall measure of organization. This measure is related to the size of the organizing units in a systematic manner, but the relation cannot be easily specified. The two extremes of the

function, however, are determined by the definition of the SO measure, and the transition from one to the other must be gradual. In one case the units of organization are individual items in the list, resulting from previous response integration (Mandler, 1954) outside the laboratory. In this case the probability of two or more items being recalled consistently together is no greater than what might be expected by chance, and the numerical value of SO is approximately the same as that of statistical subjects. On the other extreme, the whole set of items is organized into one unit, and SO is at the maximum.

Assuming a systematic relation between the size of the organizing units and the numerical value of SO, the finding that the second-order SO accounts for a larger proportion of variance in individual P scores than do estimates of higher-order measures makes good sense. Some units of organization contain only two items, and, given a flexible order of units in recall, these are not tapped by the third-order and higher-order measures of SO, while the second-order measure is sensitive to all units of two or more items. This interpretation is also consistent with the observed intercorrelations among SO measures for different lags. Although these measures are all based on identical sequences of responses, the correlations are less than perfect. This probably reflects the fact that distributions of size of organizing units vary from subject to subject.

What is the significance of the method and the preliminary findings for the question originally posed about the effects of repetition on performance? It seems that response variables such as SO, based on ordered classifications of data, might be quite useful in our attempts to shed more light on the nature of the acquisition process in verbal learning. And the present findings seem to have contributed to the attractiveness of Miller's unitization hypothesis (1956a) as a serious beginning of a useful theory of free recall verbal learning. Even though intuitively there seems little doubt that performance depends on organization, the correlational design used in the present study does not yet permit such a conclusion, and other kinds of experiments are necessary. Given a method of quantifying behavioral manifestations of the organizing process, it is quite possible to test many implications of the unitization hypothesis.

It is to be noted that the concept of repetition has been specified here rather loosely. There are at least two operationally distinguishable phases which should be studied separately, presentation of material and test for recall. A conceptual and experimental clarification of this important concept is clearly needed.

The observation of common recall sequences among different subjects and of increased stereotypy of such intersubject organization under conditions of practice very strongly suggests, or rather confirms the expectation, that sources of organization are discovered by subjects in the material, rather than invented idiosyncratically. Many task variables, therefore, are expected to influence both subjective organization and recall of the material, and a systematic exploration of these relations constitutes an important part of the research program designed to evaluate the tenability of the unitization hypothesis.

SUMMARY

A method for examining and quantifying sequential dependencies in the subjects' free recall of words on successive recall trials has been presented. Subjective organization as a depend-

ent variable was defined in terms of the subject's tendency to recall items in the same order on different trials in the absence of any experimentally manipulated sequential organization among items in the stimulus list. In a preliminary experiment it was found that the subject's recall behavior manifests such subjective organization, that this organization increases with repetition, and that there is a positive correlation between organization and performance. These data were discussed with reference to the problem of the role of repetition in free recall verbal learning.

REFERENCES

BOLLES, R. C. The effect of altering the middle of the list during serial learning. *Amer. J. Psychol.*, 1959, **72**, 577–580.

BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, 1953, **49**, 229–240.

BOUSFIELD, W. A., & COHEN, B. H. The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates. *J. Psychol.*, 1953, **36**, 67–81.

BOUSFIELD, W. A., & COHEN, B. H. The occurrence of clustering in the recall of randomly arranged words of different frequencies-of-usage. *J. gen. Psychol.*, 1955, **52**, 83–95.

BOUSFIELD, W. A., COHEN, B. H., & WHITMARSH, G. A. Associative clustering in the recall of words of different taxonomic frequencies of occurrence. *Psychol. Rep.*, 1958, **4**, 39–44.

COFER, C. N. A study of clustering in free recall based on synonyms. *J. gen. Psychol.*, 1959, **60**, 3–10.

COHEN, B. H., & BOUSFIELD, W. A. The effects of a dual-level stimulus-word list on the occurrence of clustering in recall. *J. gen. Psychol.*, 1956, **55**, 51–58.

DEESE, J. On the prediction of occurrence of particular verbal intrusions in immediate recall. *J. exp. Psychol.*, 1959, **58**, 17–22.

JENKINS, J. J., MINK, W. D., & RUSSELL, W. A. Associative clustering as a function of verbal association strength. *Psychol. Rep.*, 1958, **4**, 127–136.

JENKINS, J. J., & RUSSELL, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47**, 818–821.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MILLER, G. A. What is information measurement? *Amer. Psychologist*, 1953, **8**, 3–11.

MILLER, G. A. Note on the bias of information estimates. In H. Quastler (Ed.), *Information theory in psychology*. Glencoe, Ill.: Free Press, 1955.

MILLER, G. A. Human memory and the storage of information. *IRE Trans. inform. Theory*, 1956, **2**, 129–137. (a)

MILLER, G. A. Information and memory. *Scient. American*, 1956, **195**, 42–46. (b)

MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 1956, **63**, 81–96. (c)

MILLER, G. A., & FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.

MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. *Plans and the structure of behavior*. New York: Holt, 1960.

ROTHKOPF, E. Z., & COKE, E. U. The prediction of free recall from word association measures. *J. exp. Psychol.*, 1961, **62**, 433–438.

SAKODA, J. M. Individual differences in correlation between clustering and recall of meaningful words. *J. gen. Psychol.*, 1956, **54**, 183–190.

SHANNON, C., & WEAVER, W. *The mathematical theory of communication*. Urbana: Univer. Illinois Press, 1949.

TULVING, E., & THORNTON, G. B. Interaction between proaction and retroaction in short-term retention. *Canad. J. Psychol.*, 1959, **13**, 255–265.

(Received November 28, 1960)

Psychological Review
1962, Vol. 69, No. 4, 355–365

THE ACQUISITION OF KNOWLEDGE[1]

ROBERT M. GAGNÉ

Princeton University

The growing interest in autoinstructional devices and their component learning programs has had the effect of focusing attention on what may be called "productive learning." By this phrase is meant the kind of change in human behavior which permits the individual to perform successfully on an entire *class* of specific tasks, rather than simply on one member of the class. Self-instructional programs are designed to ensure the acquisition of capabilities of performing classes of tasks implied by names like "binary numbers," "musical notation," "solving linear equations" rather than tasks requiring the reproduction of particular responses.

When viewed in this manner, learning programming is not seen simply as a technological development incorporating previously established learning principles, but rather as one particular form of the ordering of stimulus and response events designed to bring about productive learning. It should be possible to study such learning, and the conditions which affect it, by the use of any of a variety of teaching machines, although there are few studies of this sort in the current literature (cf. Lumsdaine & Glaser, 1960). In the laboratory, the usual form taken by studies of productive learning has been primarily that of the effects of instructions and pretraining on problem solving (e.g., Hilgard, Irvine, & Whipple, 1953; Katona, 1940; Maltzman, Eisman, Brooks, & Smith, 1956).

When an individual is subjected to the situation represented by a learning program, his performance may change, and the experimenter then infers that he has acquired a new capability. It would not be adequate to say merely that he has acquired new "responses," since one cannot identify the specific responses involved. (Adding fractions, for example, could be represented by any of an infinite number of distinguishable stimulus situations, and an equal number of responses.) Since we need to have a term by means of which to refer to what is acquired as a result of responding correctly to a learning program, we may as well use the term "knowledge." By definition, "knowledge" is that inferred capability which makes possible the successful performance of a *class of tasks* that could not be performed before the learning was undertaken.

Some Initial Observations

In a previous study of programmed learning (Gagné & Brown, 1961) several kinds of learning programs were used in the attempt to establish the performance, in high school boys, of deriving formulas for the sum of n terms in a number series. Additional observations with this material led us to the following formulation: In productive learning, we are dealing with two major categories of variables. The first of these is knowledge, that is, the capabilities the individual possesses at any given stage in the learning; while the second is instructions,

[1] This study was made possible in part by funds granted by the Carnegie Corporation of New York. The opinions expressed are those of the author, and do not necessarily reflect the views of that Corporation.

the content of the communications presented within the frames of a learning program.

In considering further the knowledge category, it has been found possible to identify this class of variable more comprehensively in the following way: Beginning with the final task, the question is asked, What kind of capability would an individual have to possess if he were able to perform this task successfully, were we to give him only instructions? The answer to this question, it turns out, identifies a new class of task which appears to have several important characteristics. Although it is conceived as an internal "disposition," it is directly measurable as a performance. Yet it is *not the same* performance as the final task from which it was derived. It is in some sense *simpler*, and it is also *more general*. In other words, it appears that what we have defined by this procedure is an entity of "subordinate knowledge" which is essential to the performance of the more specific final task.

Having done this, it was natural to think next of repeating the procedure with this newly defined entity (task). What would the individual have to know in order to be capable of doing *this* task without undertaking any learning, but given only some instructions? This time it seemed evident that there were two entities of subordinate knowledge which combined in support of the task. Continuing to follow this procedure, we found that what we were defining was a *hierarchy* of subordinate knowledges, growing increasingly "simple," and at the same time increasingly general as the defining process continued.

By means of this systematic analysis, it was possible to identify nine separate entities of subordinate knowledge, arranged in hierarchical fashion (see Figure 1). Generally stated, our hypothesis was that (*a*) no individual could perform the final task without having these subordinate capabilities (i.e., without being able to perform these simpler and more general tasks); and (*b*) that any superordinate task in the hierarchy could be performed by an individual provided suitable instructions were given, and provided the relevant subordinate knowledges could be recalled by him.

It may be noted that there are some possible resemblances between the entities of such a knowledge hierarchy and the hypothetical constructs described by three other writers. First are the habit-family hierarchies of Maltzman (1955), which are conceived to mediate problem solving, and are aroused by instructions (Maltzman et al., 1956). The second are the "organizations" proposed by Katona (1940), which are considered to be combined by the learner into new knowledge after receiving certain kinds of instructions, without repetitive practice. The third is Harlow's (1949) concept of learning set. Harlow's monkeys acquired a general capability of successfully performing a class of tasks, such as oddity problems, and accordingly are said to have acquired a learning set. There is also the suggestion in one of Harlow's (Harlow & Harlow, 1949) reports that there may be a hierarchical arrangement of tasks more complex than oddity problems which monkeys can successfully perform. Since we think it important to imply a continuity between the relatively complex performances described here and the simpler ones performed by monkeys, we are inclined to refer to these subordinate capabilities as "learning sets."

REQUIREMENTS OF THEORY

If there is to be a theory of productive learning, it evidently must deal with the independent variables that can

be identified in the two major categories of instructions and subordinate capabilities, as well as with their interactions, in bringing about changes in human performance.

Instructions

Within a learning program, instructions generally take the form of sentences which communicate something to the learner. It seems possible to think of such "communication" as being carried out with animals lower than man, by means of quite a different set of experimental operations. Because of these communications, the human learner progresses from a point in the learning sequence at which he can perform one set of tasks to a point at which he achieves, for the first time, a higher level learning set (class of tasks). What functions must a theory of knowledge acquisition account for, if it is to encompass the effects of instructions? The following paragraphs will attempt to describe these functions, not necessarily in order of importance.

First, instructions make it possible for the learner to *identify the required terminal performance* (for any given learning set). In educational terms, it might be said that they "define the goal." For example, if the task is adding fractions, it may be necessary for the learner to identify $15\frac{3}{4}$ as an adequate answer, and $6\frac{3}{4}$ as an inadequate one.

Second, instructions bring about proper *identifications of the elements* of the stimulus situation. For example, suppose that problems are to be presented using the word "fraction." The learner must be able to identify $\frac{2}{5}$ as a fraction and .4 as not a fraction. Or, he may have to identify Σ as "sum of," and n as "number." Usually, instructions establish such identifications in a very few repetitions, and sometimes in a single trial. If there are many of them, differentiation may require several repetitions involving contrasting feedback for right and wrong responses.

A third function of instructions is to establish *high recallability* of learning sets. The most obviously manipulable way to do this is by repetition. However it should be noted that repetition has a particular meaning in this context. It is not exact repetition of a stimulus situation (as in reproductive learning), but rather the presentation of additional examples of a class of tasks. Typically, within a learning program, a task representing a particular learning set is achieved once, for the first time. This may then be followed by instructions which present one or more additional examples of this same class of task. "Variety" in such repetition (meaning variety in the stimulus context) may be an important subvariable in affecting recallability. Instructions having the function of establishing high recallability for learning sets may demand "recall," as in the instances cited, or they may on other occasions attempt to achieve this effect by "recognition" (i.e., not requiring the learner to produce an answer).

The fourth function of instructions is perhaps the most interesting from the standpoint of the questions it raises for research. This is the "guidance of thinking," concerning whose operation there is only a small amount of evidence (cf. Duncan, 1959). Once the subordinate learning sets have been recalled, instructions are used to promote their application to (or perhaps "integration into") the performance of a task that is entirely new so far as the learner is concerned. At a minimum, this function of instructions may be provided by a statement like "Now put these ideas together to solve this problem;" possibly this amounts to an attempt to establish a *set*. Beyond this, thinking may be

guided by suggestions which progressively limit the range of hypotheses entertained by the learner, in such a way as to decrease the number of incorrect solutions he considers (cf. Gagné & Brown, 1961; Katona, 1940). Within a typical learning program, guidance of thinking is employed after identification of terminal performance and of stimulus elements have been completed, and after high recallability of relevant learning sets has been ensured. In common sense terms, the purpose of these instructions is to suggest to the learner "how to approach the solution of a new task" without, however, "telling him the answer."

Obviously, much more is needed to be known about the effects of this variable, if indeed it is a single variable. Initially, it might be noted that guidance of thinking can vary in *amount;* that is, one can design a set of instructions which say no more than "now do this new task" (a minimal amount); or, at the other end of the scale, a set of instructions which in effect suggest a step-by-step procedure for using previously acquired learning sets in a new situation.

Subordinate Capabilities: Learning Sets

When one begins with the performance of a particular class of tasks as a criterion of terminal behavior, it is possible to identify the subordinate learning sets required by means of the procedure previously described. The question may be stated more exactly as, "What would the individual have to be able to do in order that he can attain successful performance on this task, provided he is given only instructions?" This question is then applied successively to the subordinate classes of tasks identified by the answer. "What he would have to be able to do" is in each case one or more performances which constitute the denotative definitions of learning sets for particular classes of tasks, and totally for the entire knowledge hierarchy.

A theory of knowledge acquisition must propose some manner of functioning for the learning sets in a hierarchy. A good possibility seems to be that they are mediators of positive transfer from lower-level learning sets to higher-level tasks. The hypothesis is proposed that specific transfer from one learning set to another standing above it in the hierarchy will be zero if the lower one cannot be recalled, and will range up to 100% if it can be.

In narrative form, the action of the two classes of variables in the acquisition of knowledge is conceived in the following way. A human learner begins the acquisition of the capability of performing a particular class of tasks with an individual array of relevant learning sets, previously acquired. He then acquires new learning sets at progressively higher levels of the knowledge hierarchy until the final class of tasks is achieved. Attaining each new learning set depends upon a process of positive transfer, which is dependent upon (*a*) the recall of relevant subordinate learning sets, and upon (*b*) the effects of instructions.

EXPERIMENTAL PREDICTIONS AND RESULTS[2]

Using the procedure described, we derived the knowledge hierarchy depicted in Figure 1 for the final task of "deriving formulas for the sum of *n* terms in number series."

As mentioned previously, it contained nine hypothesized learning sets. (The final row of circled entities will

[2] The author is grateful to Bert Zippel, Jr., for assistance in the preparation of learning program materials and in the collection of a portion of the data.

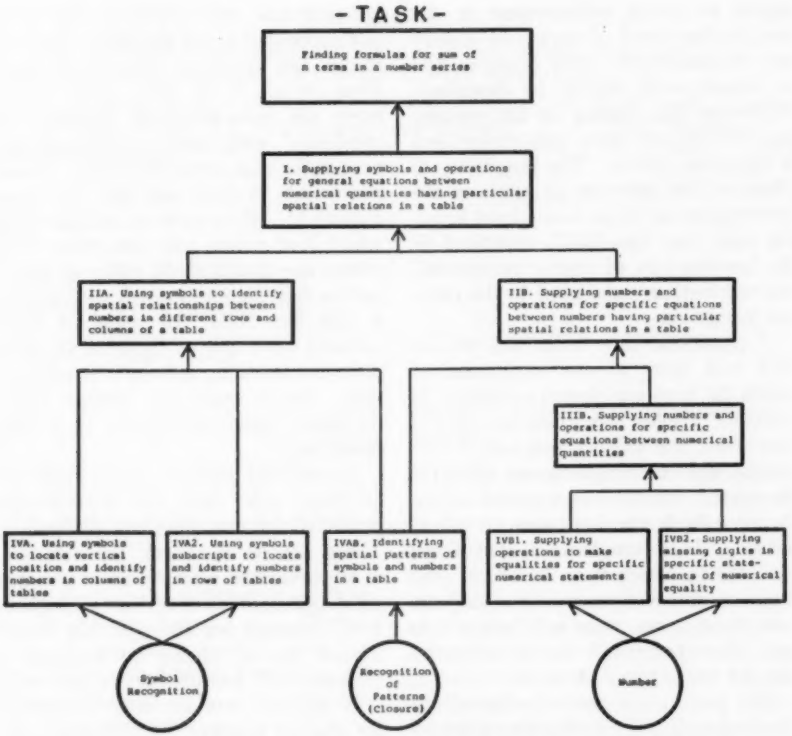
-TASK-
Finding formulas for sum of n terms in a number series
I. Supplying symbols and operations for general equations between numerical quantities having particular spatial relations in a table
IIA. Using symbols to identify spatial relationships between numbers in different rows and columns of a table
IIB. Supplying numbers and operations for specific equations between numbers having particular spatial relations in a table
IIIB. Supplying numbers and operations for specific equations between numerical quantities
IVA. Using symbols to locate vertical position and identify numbers in columns of tables
IVA2. Using symbols subscripts to locate and identify numbers in rows of tables
IVAB. Identifying spatial patterns of symbols and numbers in a table
IVB1. Supplying operations to make equalities for specific numerical statements
IVB2. Supplying missing digits in specific statements of numerical equality
Symbol Recognition
Recognition of Patterns (Closure)
Number

FIG. 1. Hierarchy of knowledge for the task of finding formulas for the for the sum of n terms in a number series.

be discussed later.) Each of these subordinate knowledges can be represented as a class of task to be performed.

Measuring Initial Patterns of Learning Sets

It is predicted that the presence of different patterns of learning sets can be determined for individuals who are unable to perform a final task such as the one under consideration. To test this, we administered a series of test items to a number of ninth-grade boys. These items were presented on $4'' \times 6''$ cards, and the answers were written on specially prepared answer sheets. This particular method was used in order to make testing continuous with the administration of a learning program to be described hereafter. Each test item was carefully prepared to include instructions having the function of identification of terminal performance and of elements of the stimulus situation.

Beginning with the final task, the items were arranged to be presented in the order I, IIA, IVA, IVA2, IVAB, IIB, IIIB, IVB1, and IVB2. For any given subject, the sequence of testing temporarily stopped at the level at which successful performance was first reached, and a learning program de-

signed to foster achievement at the next higher level (previously failed) was administered. This program and its results will shortly be described. Following this, testing on the remaining learning set tasks was undertaken in the order given. The possibility of effects of the learning program on the performance of these lower-level learning sets (not specifically practiced in the learning) is of course recognized, but not further considered in the present discussion.

A particular time limit was set for each test item, at the expiration of which the item was scored as failed. If a wrong answer was given before this time limit, the subject was told it was wrong, and encouraged to try again; if the correct answer was supplied within the time limit, the item was scored as passed. It is emphasized that these time limits, which were based on preliminary observations on other subjects with these tasks, were *not* designed to put "time pressure" on the subjects, nor did they appear to do so.

The patterns of success achieved on the final task and all subordinate learning set tasks, by all seven subjects, are shown in Table 1. The subjects have been arranged in accordance with their degree of success with all tasks, beginning with one who failed the final task but succeeded at all the rest. Several things are apparent from these data. First of all, it is quite evident that there are quite different "patterns of capability" with which individuals approach the task set by the study. Some are unable to do a task like IIA (see Figure 1), others to do a task like IIB which is of course quite different. Still others are unable to do either of these, and in fact cannot perform successfully a task like IIIB. All seven of these subjects were able to perform IV-level tasks successfully, although in preliminary observations on similar tasks we found some ninth-grade boys who could not.

Second, the patterns of pass and fail on these tasks have the relationships predicted by the previous discussion. There are no instances, for example, of an individual who is able to perform what has been identified as a "higher-level" learning set, and who then shows himself to be unable to perform a "lower-level" learning set related to it.

If learning sets are indeed essential for positive transfer, the following consequences should ensue:

1. If a higher-level learning set is passed (+), *all* related lower-level tasks must have been passed (+).

TABLE 1

PATTERN OF SUCCESS ON LEARNING SET TASKS RELATED TO THE FINAL NUMBER SERIES TASK FOR SEVEN NINTH GRADE BOYS

Subject	Task									
	Final	I	IIA	IIB	IIIB	IVA1	IVA2	IVAB	IVB1	IVB2
WW	−	+	+	+	+	+	+	+	+	+
WC	−	+	+	+	+	+	+	+	+	+
PM	−	−	−	+	+	+	+	+	+	+
GR	−	−	−	+	+	+	+	+	+	+
DJ	−	−	−	−	+	+	+	+	+	+
JR	−	−	−	−	−	+	+	+	+	+
RH	−	−	−	−	−	+	+	+	+	+

Note.—+ = Pass; − = Fail.

TABLE 2

PASS-FAIL RELATIONSHIP BETWEEN RELATED ADJACENT HIGHER- AND LOWER-LEVEL LEARNING SETS FOR A GROUP OF SEVEN NINTH-GRADE BOYS

Relationship examined	Number of cases with relationship				Test of relationships	
	Higher + Lower +	Higher − Lower −	Higher + Lower −	Higher − Lower +	N (1+2+3)	Proportion (1+2)/(1+2+3)
Final Task: I	0	6	0	1	6	1.00
I: IIA, IIB	1	5	0	1	6	1.00
IIA: IVA1, IVA2, IVAB	2	0	0	5	2	1.00
IIB: IIIB	3	2	0	2	5	1.00
IIIB: IVAB, IVB1, IVB2	5	0	0	2	5	1.00

Note.—+ = Pass; − = Fail.

2. If *one or more* lower-level tasks have been failed (−), the related higher-level tasks must be failed (−).

3. If a higher-level task is passed (+), *no* related lower-level tasks must have been failed (−).

4. If a higher-level task has been failed (−), related lower-level tasks may have been passed (+). The absence of positive transfer in this case would be attributable to a deficiency in instructions, and does not contradict the notion that lower-level sets are essential to the achievement of higher-level ones.

The relationships found to exist in these seven subjects are summarized in Table 2, where each possible higher–lower-level task relationship possible of testing is listed in the left-hand column. It will be noted that there are several relationships of the type higher (−), lower (+), as listed in Column 5. These provided no test of the hypothesis regarding hierarchical relations among learning sets. The instances in the remaining columns do, however. The ++ and −− instances are verifying, whereas +− instances would be nonverifying. As the final column indicates, the percentage of verifying instances is in all cases 100%.

Effects of Learning Program Administration

If the characteristics of instructions as previously described are correct, it should be possible to construct a learning program which can be begun for each individual at the point of his lowest successful learning set achievement, and bring him to successful achievement of the final class of tasks. Briefly, its method should be to include frames which have the functions of (*a*) insuring high recallability of relevant learning sets on which achievement has been demonstrated; (*b*) making possible identifications of expected performance and of new stimuli, for each newly presented task; and (*c*) guiding thinking so as to suggest proper directions for hypotheses associating subordinate learning sets with each new one.

A program of this sort was administered to each of the seven ninth-grade boys, beginning at the level at which he first attained success on learning set tasks (Table 1). This was done by means of a simple teaching machine consisting of a visible card file clipped to a board mounted at a 40° angle to the learner's table, and containing material typed on 4″ × 6″ cards. He wrote his answer to successive frames on a numbered answer sheet, then

flipped over the card to see the correct answer on the back. He was instructed that if his answer was wrong, he should flip the card back, and read the frame again until he could "see" what the right answer was.

After completing the instructional portion of the program for each learning set, the learner was again presented with the identical test-item problem he had tried previously and failed. If he was now able to do it correctly, he was given five additional items of the same sort to perform, and then taken on by instructions to another learning set in either a coordinate or higher-level position in the hierarchy. This process was continued through the performance of the final task.

The data collected in this way yield pass-fail scores on each test item (representing a particular member of a class of tasks) *before* the administration of the learning program, and similar scores on the same item *after* learning. It is recognized that for certain experimental purposes, one would wish to have a different, matched, task for the test given after learning, to control for the effects of "acquaintance" during the first test. Since this study had an exploratory character, such a control was not used this time. However, it should be clearly understood that the first experience with these test items in question, for these subjects, involved only activity terminating in failure to achieve solution. No information about the correct solutions was given.

A striking number of instances of success in achieving correct solutions to learning set tasks was found following learning as compared with before. These results are summarized in Table 3. Although for learning set IIB the percentage of success was only 50% (with two cases), there were two learning sets for which 100% success was achieved, and the percentage for all instances combined was 86%. These results provide additional evidence compatible with the idea of the knowledge hierarchy.

TABLE 3

NUMBER OF INSTANCES OF PASSING AND FAILING FINAL TASK AND SUBORDINATE LEARNING SET TASKS BEFORE AND AFTER ADMINISTRATION OF AN ADAPTIVE LEARNING PROGRAM, IN A GROUP OF SEVEN NINTH-GRADE BOYS

Task	Number failing before learning	Number of these passing after learning	Percentage success
Final Task	7	6	86
I	5	4	80
II	5	5	100
IIB	2	1	50
IIIB	2	2	100
Total	21	18	86

The learner in such a program does not "practice the final task"; he acquires specifically identified capabilities in a specified order. In as many as six out of seven cases, we were able by this means to bring learners from various levels of competence all the way to final task achievement. (It is perhaps important that the exception was JR, one of two who had most to learn.) Of course, it must be recognized that two separable causes contribute to the effects of the learning program in this study: (*a*) the correctness of the learning set analysis; and (*b*) the specific effectiveness of the instructions contained in the learning program.

IMPLICATIONS FOR INDIVIDUAL DIFFERENCES MEASUREMENT

It is evident that learning sets, as conceived in this paper, operate as "individual differences" variables, which, when suitably manipulated, also become "experimental" variables. There

are some additional implications which need to be pointed out regarding the functioning of learning sets in the determination of measured individual differences.

As the process of identification of subordinate learning sets is progressively continued, one arrives at some learning sets which are very simple and general, and likely to be widespread within the population of learners for which the task is designed. Consider, for example, learning set IVB (Figure 1), which is represented by a task such as $4 \times 2 = 5+?$ If one makes a further analysis to identify a subordinate learning set for this task, the answer appears to be, "adding, subtracting, multplying, and dividing one- and two-place numbers." It is interesting to note that this is exactly the task provided by a set of factor reference tests (French, 1954) called Number. In a similar manner, the other two circled entities in the last row of Figure 1 were identified. One is Symbol Recognition (called Associative Memory by the factor researchers), and Recognition of Patterns (called Flexibility of Closure). The implication is, therefore, that these simplest tasks, identified by factor analysis techniques as common to a great variety of human performances, also function as learning sets.

The hypothesis has been proposed that learning sets mediate positive transfer to higher-level tasks. Very often, if not usually, the measurement of transfer of training implies that a second task is learned *more* rapidly when preceded by the learning of an initial task, than when not so preceded. Accordingly, it seems necessary to distinguish between expected correlations of these basic factors (at the bottom of the hierarchy) with rate of attainment of higher-level learning sets on the one hand, and correlations of these same factors with achievement of higher-level learning sets on the other.

The implications of this line of reasoning would seem to be somewhat as follows: Factors which are found by the kind of psychological analysis previously described to lie at the bottom of the knowledge hierarchy should exhibit certain predictable patterns of correlation with higher-level learning sets. They should correlate most highly with rate of attainment of the learning sets in the next higher level to which they are related, and progressively less as one progresses upwards in the hierarchy. The reason for this is simply that the rate of attainment of learning sets in a hierarchy comes to depend to an increasing extent on the learning sets which have just previously been acquired and accordingly to a decreasing extent upon a basic factor or ability. Some analogy may be drawn here with the findings of Fleishman and Hempel (1954) on motor tasks.

The expected relationships between factor test scores and achievement scores (passing or failing learning sets) throughout such hierarchies seem to require a somewhat more complex derivation. First of all, such relationships will depend upon the effectiveness of a learning program, or perhaps on the effectiveness of previous learning. If the learning program is perfectly effective, for example, and if differences in rate of attainment are ignored, everyone will pass all the learning set tasks, and the variance will accordingly be reduced to zero. Under these circumstances, then, one may expect all correlations with basic factors to be zero. However, one must consider the case in which the learning program is not perfectly effective. In such a case, the probability that an individual will acquire a new learning set, as opposed to not acquiring it, will presumably be increased to the extent

that he scores high on tests of related basic abilities. If one continues to collect scores on learning set tasks of both successful achievers and those who fail, the result will presumably be an increasing degree of correlation between basic ability scores and learning set tasks as one progresses upwards in the hierarchy. The reason for this is that the size of the correlation comes to depend more and more upon variance contributed by those individuals who are successful, and less and less on that contributed by those who effectively "drop out."

The difference in expectation between the increasing pattern of correlation with achievement scores, and the decreasing pattern with measures of rate of attainment, is considered to be of rather general importance for the area of individual differences measurement. Confirmatory results have been obtained in a recent study (Gagné & Paradise, 1961) concerned with the class of tasks "solving linear algebraic equations."

DISCUSSION

The general view of productive learning implied in this paper is that it is a matter of transfer of training from component learning sets to a new activity which incorporates these previously acquired capabilities. This new activity so produced is qualitatively different from the tasks which correspond to the "old" learning sets; that is, it must be described by a different set of operations, rather than simply being "more difficult." The characteristics of tasks which make achievement of one class of task the required precursor of achievement in another, and not vice versa, are yet to be discovered. Sufficient examples exist of this phenomenon to convince one of its reality (Gagné, Mayor, Garstens, & Paradise, 1962; Gagné & Paradise, 1961). What remains to be done, presumably, is to begin with extremely simple levels of task, such as discriminations, and investigate transfer of training to tasks of greater and greater degrees of complexity, or perhaps abstractness, thus determining the dimensions which make transfer possible.

The path to research on the characteristics of instructions appears more straightforward, at least at first glance. The establishment of identifications is a matter which has been investigated extensively with the use of paired associates. The employment of instructions for this purpose may need to take into consideration the necessity for learning differentiations among the stimulus items to be identified, as well as other variables suggested by verbal learning studies. The function of inducing high recallability would seem to be a matter related to repetition of learning set tasks, and may in addition be related to time variables such as those involved in distribution of practice. As for guidance of thinking, the distinguishing of this function from others performed by instructions should at least make possible the design of more highly analytical studies than have been possible in the past.

In the meantime, the approach employed in the experiment reported here, of proceeding backwards by analysis of an already existing task, has much to recommend it as a way of understanding the learning of school subjects like mathematics and science, and perhaps others also. Naturally, every human task yields a different hierarchy of learning sets when this method of analysis is applied. Often, the relationship of higher to lower learning sets is more complex than that exhibited in Figure 1. It should be possible, beginning with any existing class of tasks, to investigate the effects of various instruction variables

within the framework of suitably designed learning programs.

The major methodological implication of this paper is to the effect that investigations of productive learning must deal intensively with the kinds of variables usually classified as "individual differences." One cannot depend upon a measurement of *general* proficiency or aptitude to reveal much of the important variability in the capabilities people bring with them to a given task. Consider, for example, the seven ninth-grade boys in our study. Each of them had "had" algebra, and each of them had "had" arithmetic. There was no particularly striking relationship between their ultimate performance and their previous grades in algebra (although there is no doubt some correlation), nor between this performance and "general intelligence." But the measurement of their learning sets, as illustrated in Table 1, revealed a great deal about how they would behave when confronted with the learning program and the final task. For some, instructions had to begin, in effect, "lower down" than for others. Some could do Task 1 right away, while others could not, but could do it equally well provided they learned other things first. The methodological point is simply this: if one wants to investigate the effects of an experimental treatment on the behavior of individuals or groups who start from the same point, he would be well advised to measure and map out for each individual the learning sets relevant to the experimental task. In this way he can have some assurance of the extent to which his subjects are equivalent.

REFERENCES

DUNCAN, C. P. Recent research on human problem solving. *Psychol. Bull.*, 1959, **56**, 397–429.

FLEISHMAN, E. A., & HEMPEL, W. E. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **19**, 239–252.

FRENCH, J. W. (Ed.) *Kit of selected tests for reference aptitude and achievement factors.* Princeton, N. J.: Educational Testing Service, 1954.

GAGNÉ, R. M., & BROWN, L. T. Some factors in the programing of conceptual learning. *J. exp. Psychol.*, 1961, **62**, 313–321.

GAGNÉ, R. M., MAYOR, J. R., GARSTENS, H. L., & PARADISE, N. E. Factors in acquiring knowledge of a mathematical task. *Psychol. Monogr.*, 1962, **76**(7, Whole No. 526.

GAGNÉ, R. M., & PARADISE, N. E. Abilities and learning sets in knowledge acquisition. *Psychol. Monogr.*, 1961, **75**(14, Whole No. 518).

HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

HARLOW, H. F., & HARLOW, M. K. Learning to think. *Scient. American*, 1949, **181**, 36–39.

HILGARD, E. R., IRVINE, R. P., & WHIPPLE, J. E. Rote memorization, understanding, and transfer: An extension of Katona's card-trick experiments. *J. exp. Psychol.*, 1953, **46**, 288–292.

KATONA, G. *Organizing and memorizing.* New York: Columbia Univer. Press, 1940.

LUMSDAINE, A. A., & GLASER, R. *Teaching machines and programmed learning.* Washington, D. C.: National Education Association, 1960.

MALTZMAN, I. Thinking: From a behavioristic point of view. *Psychol. Rev.*, 1955, **62**, 275–286.

MALTZMAN, I., EISMAN, E., BROOKS, L. O., & SMITH, W. M. Task instructions for anagrams following different task instructions and training. *J. exp. Psychol.*, 1956, **51**, 418–420.

(Early publication received August 8, 1961)

Psychological Review
1962, Vol. 69, No. 4, 366–378

ON THE RELATION BETWEEN LOGIC AND THINKING

MARY HENLE[1]

Graduate Faculty of Political and Social Science,
New School for Social Research, New York

The question of whether logic is descriptive of the thinking process, or whether its relation to thinking is normative only, seems to be easily answered. Our reasoning does not, for example, ordinarily follow the syllogistic form; and we do fall into contradictions. On the other hand, logic unquestionably provides criteria by which the validity of reasoning may be evaluated. Logical forms thus do not describe actual thinking, but are concerned with the ideal, with "how we ought to think." And yet a problem seems to be concealed beneath this easy solution.

It is interesting to note that a number of the older writers on logic regarded their discipline as the science of thought. It will not be possible here to survey the vast literature, some of which bears only by implication on our problem, but some examples will illustrate this position. Cohen and Nagel (1934) summarize the older view: "An old tradition defines logic as the science of the laws of thought" (p. 18). Kant (1885) holds that "logic is a science of the necessary laws of thought, without which no employment of the understanding and the reason takes place" (p. 3). Psychology, on the other hand, supplies only the contingent, not the necessary, rules of thought. John Stuart Mill (1874) likewise views logic as comprising "the science of reasoning, as well as an art, founded on that science" (p. 18). He continues:

> Whatever has at any time been concluded justly, whatever knowledge has been acquired otherwise than by immediate intuition, depended on the observance of the laws which it is the province of logic to investigate. If the conclusions are just, and the knowledge real, those laws, whether known or not, have been observed (p. 22).

Boole (1854), as a final instance, regards "the laws of the symbols of Logic" as "deducible from a consideration of the operations of the mind in reasoning" (pp. 45–46). Boole deduces the law of contradiction, for example, from a fundamental law of thought (pp. 49–51).

More recent writers, on the other hand, have tended to reject the view that the laws of logic are those of the human understanding. Again, only a few illustrations will be given. Cohen (1944) remarks:

> That the laws of logic are not the universal laws according to which we do actually think is conclusively shown, not only by the most elementary observation or introspection, but by the very existence of fallacies (pp. 2–3).

A similar point is made by Nagel (1956):

> Little need be said in refutation of the view that logical principles formulate the "inherent necessities of thought" and are generalized descriptions of the operations of minds. Surely the actual occurrence of beliefs in logically incompatible propositions makes nonsense of the claim that the principle of noncontradiction expresses a universal fact of psychology (p. 66).

Regarding the work of Boole referred to above (*An Investigation of the Laws of Thought*), Cohen and Nagel (1934) comment: "The title is a misnomer"

[1] This paper was written when the author was a Fellow of the John Simon Guggenheim Memorial Foundation.

(p. 112). A more extreme statement of this point of view is by Bertrand Russell (1904):

> Throughout logic and mathematics, the existence of the human or any other mind is totally irrelevant; mental processes are studied by means of logic, but the subject-matter of logic does not presuppose mental processes, and would be equally true if there were no mental processes (p. 812).

Schiller (1930), too, holds that syllogistic reasoning "has nothing whatever to do with actual reasoning, and can make nothing of it" (p. 282). He describes the "Laws of Thought" (laws of identity, contradiction, and excluded middle) as "verbal conventions" (p. 251).

The changed point of view with regard to the relation between logical principles and the laws of thought seems to be a function of an altered intellectual climate rather than of any fundamental discoveries about the nature of reasoning. It thus seems worthwhile to reopen the question, the more so since it has implications for a number of central issues in the psychology of thinking. First, however, it may be of interest to examine a few more forms in which the present question has been raised.

Discussion of the figures of the syllogism has at times centered on their relevance to actual thinking. J. N. Keynes (1887) cites several writers who reject the fourth figure because they hold that we do not actually reason in it (pp. 230–231). Kant (1885), indeed, finds all but the first figure both useless and false; the fourth in particular he calls unnatural (pp. 84–90). Keynes, on the other hand, argues for the admission of Figure 4 on the same grounds, namely its relevance to actual thinking: "It is not actually in frequent use, but reasonings may sometimes not unnaturally fall into it" (p. 232).

Psychologists investigating reasoning processes have tended to underemphasize the role of logic in the thinking of their subjects. To illustrate, Bruner, Goodnow, and Austin (1956) suggest that

> much of human reasoning is supported by a kind of thematic process rather than by an abstract logic. The principal feature of this thematic process is its pragmatic rather than its logical structure (p. 104).

Individuals tend to prefer "empirically reasonable propositions" to logical ones (p. 104). Morgan and Morton (1944) conclude that:

> A person is likely to accept a conclusion which expresses his convictions with little regard for the correctness or incorrectness of the inferences involved. Our evidence will indicate that the only circumstance under which we can be relatively sure that the inferences of a person will be logical is when they lead to a conclusion which he has already accepted (p. 39).

Lefford (1946) states that the principles of logical inference "are techniques which are not the common property of the unsophisticated subject" (p. 144). He goes so far as to distinguish from the logical inference

> psychological inferences which may be made by the ordinary person. . . . A psychological inference is not valid or invalid except when judged as a logical inference: psychological inference is purely a fact (p. 145).

Common to all these statements by the psychologists is the assumption that logical principles are irrelevant, if not antithetical, to much actual reasoning.[2] This conclusion is derived from the

[2] A different, but related, position is that represented by Dollard and Miller (1950), who hold that being logical is a learned drive. The child, it is argued, is punished for logical contradictions and absurdities, for illogical and contradictory plans. The result for most people is "a learned drive to make their explanations and plans seem logical" (p. 120). The implication is, of course, that without this specific training the individual's thinking would not be (or seem) logical.

high incidence of wrong inferences of subjects under test conditions, especially in the case of emotionally relevant material. A different conclusion has been drawn by Von Domarus (1944) from observation of errors in the reasoning of schizophrenic patients. This author argues, not that the reasoning of his subjects is unrelated to logic, but rather that it conforms to a logic whose laws are different from those of Aristotelian logic. This is "paralogic," which excludes the law of contradiction and "accepts identity based upon identical predicates" (p. 111). This idea has been elaborated by Arieti (1955) who sees the operation of this logic (which he calls paleologic) not only in schizophrenic and primitive thinking, as Von Domarus does, but also in dreams, in some infantile thinking, and in the transference situation in psychoanalysis.

Again it is of interest to find that the earlier writers mentioned above were equally aware of the problem of error, but viewed it in a way that was entirely compatible with their conception of logic as the science of the laws of the mind. "It is easy to see how truth is possible," writes Kant (1885, p. 44) "since in it the understanding acts according to its own essential laws." Error, however, is difficult to understand since it constitutes "a form of thought inconsistent with the understanding." Its source is thus not to be sought in the understanding itself, but rather in the "unobserved influence of the sensibility on the understanding," the sensibility being that faculty which "supplies the material for thought." Boole (1854) likewise considers that "the phaenomena of incorrect reasoning or error . . . are due to the interference of other laws with those laws of which *right* reasoning is the product" (p. 409). He reminds us that "the laws of correct inference may be violated, but they do not the less truly *exist* on this account" (p. 408). Mill (1874) is still more explicit. Discussing fallacies of ratiocination, he points out that since

> the premises are seldom formally set out, . . . it is almost always to a certain degree optional in what manner the suppressed link shall be filled up. . . . [A person] has it almost always in his power to make his syllogism good by introducing a false premise; and hence it is scarcely ever possible decidedly to affirm that any argument involves a bad syllogism (p. 560).

In the case of arguments consisting not of a single syllogism but of a chain of syllogisms, he considers the commonest fallacy of ratiocination to lie in a changing of the premises as the argument proceeds.

Two clearly contrasting alternatives thus present themselves: Is logic (or Aristotelian logic) largely irrelevant to the thinking process, or is it concerned with the laws of thinking? Since we will here be concerned only with deductive reasoning, we may reformulate the question more specifically in these terms. But since, as has so often been pointed out, the premises from which we reason are commonly not spelled out, since our inferences so frequently appear as enthymemes,[3] this fact must be taken into consideration. We may ask: If we know the premises—tacit as well as explicit—from which a person reasons, can we put the process in syllogistic form? Do the rules of the syllogism describe processes that the mind follows in deductive reasoning, even when the syllogistic form is not explicitly employed?

It has been shown above that the existence of error has been used as evidence for the irrelevance of logic to the actual thinking process. On the other hand, a different interpretation of error

[3] "A syllogism that is incompletely stated, in which one of the premises or the conclusion is tacitly present but not expressed, is called an enthymeme" (Cohen & Nagel, 1934, p. 78).

has been suggested. Since the problem of error seems to be a particularly fruitful one in which to join the issues before us, it will be taken as the context for the present discussion. Once more we may reformulate our question as follows: Do errors in deductive reasoning mean that the logical process has been violated? As Mill expresses it, does the occurrence of error mean that the syllogism is a bad one? Or can the error be accounted for otherwise? Is it possible that a process that would follow the rules of logic if it were spelled out is discernible even when the reasoning results in error?

The distinction being made here is a familiar one in the psychology of learning and thinking. Thus Koffka (1935, Ch. 12) distinguishes between learning as accomplishment and the learning processes responsible for this accomplishment. Köhler's (1927) distinction between "good errors" and stupid ones is likewise relevant. Good errors, he points out, "may, in a certain sense, be absolutely appropriate to the situation" (p. 217), although they solve the problem no more than do stupid ones. Wertheimer (1959), too, distinguishes between solutions obtained by "blind" procedures and "fine, genuine solutions." Again the difference is one of process, since in both cases the result may be the same. In the same way, in connection with the present problem, we may ask: Given contrasting results—correct solutions and errors in deductive reasoning—what can we say about the thinking processes that account for them? Are the processes necessarily different because their effects are different?

Illustrative data that bear on this issue will be presented. They were obtained from 46 graduate students of psychology who were asked to evaluate the logical adequacy of deductions presented in the context of everyday problems. Most of the subjects had no training in formal logic.[4] The material was presented under group conditions, the subjects writing out their judgments and their grounds for making them. Instructions included an explicit statement that the logical adequacy of the arguments was to be judged, not the truth of the statements.

Individual interviews were also conducted with a different set of subjects. Although the results agree generally with those obtained under group conditions, they will not be presented because of a problem already mentioned. Mill, it has been seen, pointed to the difficulty of recognizing a bad syllogism because a person can easily introduce into his incompletely stated argument a new premise that will make the inference valid. It was often difficult to decide whether the material elicited by direct questioning consisted of new premises not entering into the original inference, or whether the subject was, as intended, making clear his original understanding of the premises. Although the group data are less rich than the individual interviews, and although the products of written communication are likely to be less spontaneous reflections of the thinking process than interviews, they exclude this possibility of overlooking true fallacies; indeed, if anything, they overestimate the incidence of bad syllogisms, since the reasoning is often incompletely described in the case of apparent errors.[5]

The present data will be used only to illustrate the reasoning processes in cases of error; no quantitative results

[4] It must be recognized, of course, that the lack of formal training in logic is insufficient guarantee of the naiveté of subjects. It can only be pleaded that we have no better criterion at the present time. Whatever the informal self-education of subjects in this respect, the results to be presented cannot be regarded as the product of formal instruction in logic.

[5] A sample interview may be found in Henle (1958).

will be presented. As many authors have shown, the incidence of error in deductive reasoning depends on the form of the syllogism and its contents, as well as on instructions to the subjects. Quantitative results would have relevance only to the particular conditions studied here, whereas an inquiry into the nature of the errors obtained might be of more general interest.

Several processes may be distinguished which led to error in dealing with the presented material:

FAILURE TO ACCEPT THE LOGICAL TASK

More specifically, this means failure to distinguish between a conclusion that is logically valid and one that is factually correct or one with which the subject agrees. This source of error has already been reported by Henle and Michael (1956, p. 124).

A sample syllogism follows, along with responses in which errors occur because of failure to grasp or accept the logical task.

Syllogism 6. A group of women were discussing their household problems. Mrs. Shivers broke the ice by saying: "I'm so glad we're talking about these problems. It's so important to talk about things that are in our minds. We spend so much of our time in the kitchen that of course household problems are in our minds. So it is important to talk about them." (Does it follow that it is important to talk about them? Give your reasoning.)

Responses: "The conclusion does not follow. The women must talk about household problems because it is important to talk about their problems, not because the problem is in their minds."

"No. It is not important to talk about things that are in our minds unless they worry us, which is not the case."

"No. Just because one spends 'so much time' in the kitchen it does not necessarily follow that household problems are 'in our minds'."

It should be noted that subjects who failed to accept the logical task frequently gave correct responses that are just as irrelevant to the question of the relation of logic to the thinking process as are the errors just cited. A few examples follow:

"Yes. It could be very important to the individual doing the talking and possibly to some of those listening, because it is important for people to 'get a load off their chest,' but not for any other reason, unless in the process one or the other learns something new and of value."

"Yes. It seems obvious that problems which are in the forefront of one's mind bring more consideration to them and possibly newer aspects when they are discussed with another. Two heads may be better than one."

"Yes it does. By talking household problems, a problem can be solved or worked through."

The errors illustrated here clearly do not demonstrate an inability of the subjects to reason logically, since they have not accepted the logical task. They have evaluated the content of the conclusion, not the logical form of the argument. Richter (1957, p. 341) makes the same distinction, carrying the analysis a step farther. Apart from careless mistakes and those arising from "imperfections in the classifying operation"—i.e., from an inability to make logical deductions, he describes errors arising from "a general failure to grasp the concept of 'logical validity'" and those arising from "the specific inability to differentiate 'logical validity' from another attribute of syllogisms," namely their factual status. To apply to the present data, the factual criterion needs to be interpreted to

mean what is believed to be true or reasonable as well as what is known to be true. With Richter's distinction in mind, we may note that no subject failed completely to grasp the logical task. Still, one individual judged on the basis of belief rather than logical validity in the case of 9 out of 10 syllogisms, and several others judged in this manner with at least half of the syllogisms presented.

RESTATEMENT OF A PREMISE OR CONCLUSION SO THAT THE INTENDED MEANING IS CHANGED

In a number of cases examination of the material from which a subject reasoned showed that it differed from that which was originally presented. In such cases the validity of the argument can, of course, be judged only in relation to the syllogism actually employed, not the one intended.

1. In the case of syllogisms in which a conclusion from two particular premises was to be evaluated, the premises were occasionally restated as universals, from which the conclusion followed. The following syllogism permitted this kind of change:

Syllogism 8. Mrs. Cooke had studied home economics in college. "Youth is a time of rapid growth and great demands on energy," she said. "Many youngsters don't get enough vitamins in their daily diet. And since some vitamin deficiencies are dangerous to health, it follows that the health of many of our youngsters is being endangered by inadequate diet." (Does it follow that the health of many youngsters is being endangered by inadequate diet? Give your reasoning.)

Responses: "Yes. Youngsters (A) don't get vitamins (B). Not getting vitamins (B) dangerous to health (C). Youngsters (A) in danger of poor health (C)."

"This follows. . . . Youth doesn't get enough vitamins. Vitamins are necessary or bad health results. Therefore youth are endangered. All of these follow logically, and I believe a correct inference has been made."

2. Other changes were introduced into the premises by subjects whose reasoning was found to be correct even though their answers to the intended problem were wrong. The same syllogism (Syllogism 8) will be used to illustrate.

Responses: "There is some question about the equivalence of 'don't get enough vitamins' and a vitamin deficiency dangerous to health. If one assumes their equivalence, then the conclusion just about manages to follow."

"Correct if we assume that the youngsters are lacking those vitamins in their diet which endanger health."

"It seems to follow, assuming that the deficient vitamins are also the vital ones."

Everyday discussions permit greater scope for such changes in meaning than does the material employed here. Private meanings of terms or idiosyncratic or unconscious equations of concepts may enter in to prevent two people from drawing the same conclusion even though they start ostensibly from the same material.

OMISSION OF A PREMISE

Occasionally a subject employed only one of the presented premises. The informal manner in which the premises were set out made possible this treatment of the material. Syllogism 6 (cf. above) will be used to illustrate.

Responses: "Not correct logically. (*a*) A group of women spend much of their time in the kitchen; (*b*) thus household problems occupy their minds. . . . (*c*) Therefore it is important to talk about them. Mrs. Shivers' subjective thinking."

"Doesn't follow from the preceding statement. Because she spends a lot of time in the kitchen and household problems are on her mind, it doesn't follow that it is important to talk about them."

In a few cases, indeed, both of the intended premises were omitted by the subject. For example (Syllogism 6):

"No it doesn't follow. The fact that she spends much time in the kitchen has nothing to do with whether or not it is important to talk about the problems."

It will be seen that in all these cases in which a premise has been omitted, the subject correctly reports that the conclusion does not follow. If we disregard the intended premises and consider only the material actually used in reasoning, the subjects' deductions are seen to be correct.

SLIPPING IN OF ADDITIONAL PREMISES

Under the present conditions this device was infrequent, but it seems to be more common where premises are stated less completely, and where the issues are of more immediate practical concern, as in everyday discussion. For example, premises may be added that are so commonplace, so much taken for granted, that they escape attention.

Instances of the adding of assumptions to our material follow. Since they are readily understandable, and for the sake of brevity, the responses will be quoted without the presented syllogism.

Responses: "Without intending sophistry it must be pointed out that the implicit premise is 'crime is bad.' If this is granted, the conclusion follows that if comics contribute to influence the youth along criminal paths, they should be eliminated."

". . . If comic books are an evil influence, then they should be got rid of." Here the subject is clearly employing an implicit premise that he added to the material presented, namely "Whatever is an evil influence should be got rid of."

Again in these cases, wrong answers are obtained by correct reasoning if we consider the syllogism as the subject understood it rather than the one the investigator hoped to present.

When all the above processes have been taken into account, a considerable number of errors remain. Do these represent true fallacies? It is difficult to say since in these cases either no account of the reasoning process is supplied, or an unclear explanation is given, or else the subject merely sets out the premises and conclusion given to him without comment. It might be that many of these errors would have been found on examination to involve processes similar to the ones described above, so that no error in reasoning is involved. On the other hand, the possibility of true fallacies cannot be excluded.

In connection with the unexplained errors, it is worth noting that almost as many correct responses as errors were unexplained or unclearly explained. The absence of explanation is thus no reason to suspect a reasoning process of being fallacious.

It must also be mentioned that the processes described above as producing apparent errors although the deductive process is correct are by no means viewed as constituting a complete list. They were observed with particular materials and under particular conditions; and it is to be anticipated that further research will expand this list. Chapman and Chapman (1959), for example, have suggested several additional processes by which error is produced in syllogistic reasoning, one of which involves a misunderstanding of the task different from that described here.

Despite the limitations of the present data, they do show clearly that when subjects arrive at apparently invalid conclusions, or when they fail to spot a fallacy, they often do so because they have worked with materials different from those intended or because they have undertaken a task different from the one intended. In such cases, if we consider the materials and task as they were actually understood by individual subjects, we fail to find evidence of faulty reasoning. It must be concluded that the presence of error does not constitute evidence that the laws of logic are irrelevant to actual thinking. The data tend, rather, to support the older conception that these laws are widely discernible in the thinking process.

DISCUSSION

An examination of errors in syllogistic reasoning leads to two conclusions: (*a*) While the possibility of fallacy often cannot be excluded it is, in Mill's words, "scarcely ever possible decidedly to affirm that any argument involves a bad syllogism." (*b*) Where error occurs, it need not involve faulty reasoning, but may be a function of the individual's understanding of the task or the materials presented to him. Spinoza (1930) has stated our present conclusion in more general terms:

> When men make errors in calculation, the numbers which are in their minds are not those which are upon the paper. As far as their mind is concerned there is no error, although it seems as if there were, because we think that the numbers in their minds are those which are upon the paper. If we did not think so, we should not believe them to be in error. For example, when I lately heard a man complaining that his court had flown into one of his neighbour's fowls, I understood what he meant, and therefore did not imagine him to be in error. This is the source from which so many controversies arise—that men either do not properly explain their own thoughts, or do not properly interpret those of other people; for, in truth, when they most contradict one another, they either think the same things or something different, so that those things which they suppose to be errors and absurdities in another person are not so (from the *Ethics*, Scholium Proposition 47, Part 2).

Such a conclusion, which the present analysis supports, is of some importance if, as Mill (1874) writes, "to draw inferences has been said to be the great business of life" (p. 21). Our data suggest that the intelligent adult is not so inept at carrying on this business as several recent accounts have made him out to be.

In considering the place of logical inference in everyday life, it is necessary also to take into account what Aristotle (1945) has termed the "practical syllogism," the case in which

> the conclusion drawn from the two premisses becomes the action. For example, when you conceive that every man ought to walk and you yourself are a man, you immediately walk; or if you conceive that on a particular occasion no man ought to walk, and you yourself are a man, you immediately remain at rest. In both instances action follows unless there is some hindrance or compulsion. . . . Again, I need a covering, and a cloak is a covering, I need a cloak. What I need I ought to make; I need a cloak, I ought to make a cloak. And the conclusion "I ought to make a cloak" is an action. The action results from the beginning of the train of thought. If there is to be a cloak, such and such a thing is necessary, if this thing then something else; and one immediately acts accordingly. That the action is the conclusion is quite clear. . . . (701a).

Aristotle points out that this whole process need not be explicit.

> The mind does not stop and consider at all one of the two premisses, namely, the obvious one; for example, if walking is good for a man, one does not waste time over the premiss "I am myself a man." Hence such things as we do without calculation, we do quickly (701a).

Again, "My appetite says, I must drink; this is drink, says sensation or

imagination or thought, and one immediately drinks."

If we include, as it seems we must, the practical syllogism in the "great business of life," that of drawing inferences, it becomes difficult indeed to accept the conclusion that human beings are unable to reason logically. A couple of contemporary examples will suggest how ubiquitous is the practical syllogism in everyday life. "It is too far to walk to the Public Library; I must take a subway or bus. The Fifth Avenue bus passes the Public Library. I must take the Fifth Avenue bus." "I do not want to wear the same dress two days in succession. I wore this dress yesterday; so I do not want to wear this dress today." It is difficult to see how the individual could cope with the ordinary tasks of life if the practical syllogism embodied techniques which are not, as one author quoted above put it, "the common property of the unsophisticated subject," if, indeed, it were not a natural mode of functioning of the conscious mind.

Furthermore, if people were unable to reason logically, so that each arrived at different conclusions from the same premises, it is difficult to see how they could understand each other, follow one another's thinking, reach common decisions, and work together.

It might be argued that in the case of the practical syllogism we are dealing not with an implicit logical process, but with a learning process in which each response is cued off by the preceding one. There seem to be at least two grounds for rejecting such an interpretation. (*a*) There are cases in which the practical syllogism leads to a solution which is genuinely novel for the individual. This is surely so for some of the examples cited above, at least the first time they occur. Indeed, if the responses were habitual ones, the reasoning process would not be necessary. (*b*) Our data obtained from conventional syllogisms suggest that a logical process often occurs even in cases in which the reasoning seems fallacious. Such data can thus not be interpreted in terms of the "learned drive" postulated by Dollard and Miller (1950, p. 120) "to make . . . explanations and plans seem logical." And since they sometimes lead to novel conclusions (cf. Wertheimer, 1938), they cannot be interpreted as habitual responses. In the interests of parsimony, therefore, a common explanation for the practical syllogism and the verbal one seems preferable. An interpretation in terms of an implicit logical process would fit both kinds of case.

Two further implications of the present data may be mentioned. They suggest an approach to the problem of the influence of needs and attitudes on the reasoning process. For the problem remains: why do motivational influences so often appear to impair thinking? It is a plausible hypothesis that these influences do not distort the reasoning process, as has frequently been stated or implied—indeed that they do not act at all on the reasoning process—but rather that they affect the materials with which thinking works. Although this is not the place for a detailed analysis, one or two specific effects of this kind may be suggested. (*a*) It may be that a strong attitude toward, or emotional involvement with, particular material is in part responsible for the difficulty which many unsophisticated subjects experience in distinguishing between drawing a conclusion that is logically valid and one that is believed to be correct. Over a wide range of practical situations these two tasks are not in conflict; and the distinction is thus not an easy one. There is evidence that attitudes and

emotions may limit the ability to make distinctions (cf. Henle, 1955). If this suggestion is correct, the more personally relevant the material employed, the more difficult it will be to accept the logical task. Such an effect might go a long way toward explaining the report of a number of writers that individuals are unable to reason logically about emotionally toned material. (*b*) An attitude can select from among the possibilities that the material presents, singling out, for example, one among several possible meanings. It may be that such a process is responsible for the changes in the meaning of premises observed in our data. (For possibilities of further analysis of this kind, cf. Henle, 1955.)

That motivational processes may act on reasoning in the manner suggested here receives support from a recent study by Kopp (1960), who subjected to a similar analysis reasoning processes related to the delusions of paranoid schizophrenics. His subjects, like the present ones, made errors which were found to lie more in the premises from which the reasoning proceeded than in the actual drawing of inferences.

In this connection, it would be interesting to analyze Arieti's (1955) examples of "paleological" thinking in the terms of the present study. Unfortunately, the published data are too scant to permit such an analysis. The examples presented would be fallacies in terms of Aristotelian logic if one could assume that the premises as understood by the subject are the same as those understood by the examiner. That this is the case may be doubted in a number of instances; and the search for an archaic logic thus seems premature. Some examples will be given which seem to call for an analysis different from Arieti's.

Arieti cites the case of a woman who "needed to identify herself with the Virgin Mary because of the extreme closeness and spiritual kinship she felt for the Virgin Mary" (1955, p. 195). The reasoning process is described as follows: "The Virgin Mary was a virgin; I am a virgin; therefore, I am the Virgin Mary." In the absence of additional material, we may only speculate that the proposition "I am the Virgin Mary" was not a conclusion deduced from the stated premises, but that it belongs rather in the category of revealed truth or intuition, not in that of deductive inference. It seems likely to the present writer that the syllogism quoted above was in the nature of a justification of this intuition to the examiner rather than the grounds for belief in it. The argument itself could be evaluated only in the light of additional information.

Another kind of example of Arieti's (1955) illustrates what he considers to be the propensity of young children "to indulge in paleologic thinking":

> A girl, three years and nine months old, saw two nuns walking together, and told her mother, "Mommy, look at the twins!" She thought that the nuns were twins because they were dressed alike. The characteristic of being dressed alike, which twins often have, led to the identification with the nuns (p. 199).

Again a different interpretation suggests itself: rather than viewing it as a deduction, we may take this process more simply as an instance of an inaccurate concept. Arieti also discusses dream symbols as examples of paleologic thinking. It is particularly difficult for the present writer to view such symbols as inferences, as if they were the conclusions of processes of deductive reasoning. They seem, rather, to involve a much more immediate grasping of similarities.

It is amusing to note that Arieti (1955) regards some delusions of

identification as instances of paleologic, and describes the "formal mechanism" by which they are produced in a way that could easily be restated as a valid Aristotelian syllogism:

> If A may be identified with B because they have a common quality, it will be sufficient for me to acquire a quality of the person I want to be identified with, in order to become that person (p. 207).

Further discussion of the problems raised by Arieti must await more complete data.

The present study has implications also for another problem. To label an inference fallacious, as is clear from the analysis of this paper, is to make a statement about the results of reasoning, but not about the process that is responsible for these results. It might be a matter of considerable interest to study systematically and in detail the reasoning involved in various fallacies. It might be found that many fallacies are produced, not by faulty reasoning, but by specific changes in the material as the subject understands it. Some of the syllogisms employed here involved the fallacy of the undistributed middle; here changes of premises from particulars to universals and other changes in the meaning of premises were not uncommon. Two of the arguments presented contained illicit conversions, but no relevant data were obtained since the subjects systematically ignored the conversion and centered their attention on other parts of the material.[6] In connection with illicit conversions, the discussion of Chapman and Chapman (1959) is of interest; these authors suggest implicit assumptions and changes in meaning that may be involved in the acceptance of such conversions. Mill's (1874) observation that fallacies frequently involve changing the premises has been cited above. He analyzes several examples of the fallacy *a dicto secundum quid ad dictum simpliciter*[7] in these terms (p. 561–562). Fallacies of ambiguity seem likewise to lend themselves to the kind of analysis suggested here.[8]

In conclusion, observations have been presented which are consistent

[6] This unexpected finding cannot be attributed to an insensitivity of the subjects to logical processes. Rather the fault lies with the investigator who, in the effort to disguise the formal structure of the syllogism, included irrelevancies that were taken by many subjects as the material for evaluation.

[7] "This is committed when, in the premises, a proposition is asserted with a qualification, and the qualification lost sight of in the conclusion; or oftener, when a limitation or condition, though not asserted, is necessary to the truth of the proposition, but is forgotten when that proposition comes to be employed as a premise" (Mill, 1874, p. 561).

[8] It is interesting to note that Mill's (1874) view comes close to the hypothesis offered here. It is true that he describes fallacies of ratiocination as:

> those which have their seat in the ratiocinative or deductive part of the investigation of truth (p. 559).

However he adds that it is by the rules of the syllogism

> that a reasoner is compelled distinctly to make his election what premises he is prepared to maintain. The election made, there is generally so little difficulty in seeing whether the conclusion follows from the premises set out, that we might without much logical impropriety have merged this fourth class of fallacies in the fifth, or Fallacies of Confusion (p. 560).

The latter class includes:

> all those fallacies in which the source of error is not so much a false estimate of the probative force of known evidence, as an indistinct, indefinite, and fluctuating conception of what the evidence is (p. 563).

(This class includes errors involving ambiguity of terms, begging the question, and fallacies of irrelevance.)

with the view that the rules of the syllogism describe processes that the mind follows in deductive reasoning, even when the syllogistic form is not explicitly employed.[9] A word of caution is necessary at this point. The present observations apply to deductive reasoning only. Even within this category, a very limited number of logical forms has been studied; and the study has been concerned only with the evaluation of presented arguments, not with the construction of arguments of one's own. The investigation needs to be extended before the generality of the conclusion can be assessed. It must also be pointed out that there are many types of thinking to which logical analysis does not apply at all, for example many aspects of free association, fantasy, many creative processes. To apply logical analysis to such processes (for example, Arieti's discussion of dream symbols) is as much an error as to ignore the underlying logical structure of incompletely stated arguments.

SUMMARY

Two views of the relation of logic to thinking have been presented: the older view that logic is the science of thought, and the position more widely held at present that its relation to thinking is normative only. It has been acknowledged that ordinary thinking, more often than not, does not assume the form of the syllogism, and that errors in reasoning are frequent. But does error mean that the logical process has been violated? or can we account for it in other ways? Evidence has been presented that suggests that even where the thinking process results in error, it can often be shown that it does not violate the rules of the syllogism. Many errors were found to be accounted for not in terms of a breakdown of the deductive process itself, but rather in terms of changes in the material from which the reasoning proceeds. The problem of the role of logical processes in everyday life was discussed more generally, and some implications of this study for other problems were considered.

[9] If it could be established generally that the laws of logic are those of the human mind, this would not be to give the former the status of "mere facts" or to make them relative to particular conditions; rather it would be to recognize the requiredness in thinking (cf. Köhler, 1938, especially pp. 45, 52).

REFERENCES

ARIETI, S. *Interpretation of schizophrenia.* New York: Brunner, 1955.

ARISTOTLE. *Movement of animals.* (Trans. by E. S. Forster) Cambridge: Harvard Univer. Press, 1945.

BOOLE, G. *An investigation of the laws of thought.* London: MacMillan, 1854.

BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

CHAPMAN, L. J., & CHAPMAN, JEAN P. Atmosphere effect re-examined. *J. exp. Psychol.*, 1959, **58**, 220–226.

COHEN, M. R. *A preface to logic.* New York: Holt, 1944.

COHEN, M. R., & NAGEL, E. *An introduction to logic and scientific method.* New York: Harcourt, Brace, 1934.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

HENLE, MARY. Some effects of motivational processes on cognition. *Psychol. Rev.*, 1955, **62**, 423–432.

HENLE, MARY. The birth and death of ideas. Paper read at Colorado Symposium on Creative Thinking, Boulder, 1958.

HENLE, MARY, & MICHAEL, MIRIAM. The influence of attitudes on syllogistic reasoning. *J. soc. Psychol.*, 1956, **44**, 115–127.

KANT, I. *Introduction to logic and essay on the mistaken subtilty of the four figures.* (Trans. by T. K. Abbott) London: Longman's, Green, 1885.

KEYNES, J. N. *Studies and exercises in formal logic.* (2nd ed.) London: MacMillan, 1887.

KOFFKA, K. *Principles of gestalt psychology.* New York: Harcourt, Brace, 1935.

KÖHLER, W. *The mentality of apes.* (Trans. by Ella Winter) (2nd ed.) New York: Harcourt, Brace, 1927.

KÖHLER, W. *The place of value in a world of facts.* New York: Liveright, 1938.

KOPP, S. Deductive reasoning in paranoid schizophrenics. Unpublished doctoral dissertation, New School for Social Research, 1960.

LEFFORD, A. The influence of emotional subject matter on logical reasoning. *J. gen. Psychol.,* 1946, **34**, 127–151.

MILL, J. S. *A system of logic.* (8th ed.) New York: Harper, 1874.

MORGAN, J. J. B., & MORTON, J. T. The distortion of syllogistic reasoning produced by personal convictions. *J. soc. Psychol.,* 1944, **20**, 39–59.

NAGEL, E. *Logic without metaphysics.* Glencoe, Ill.: Free Press, 1956.

RICHTER, M. N., JR. The theoretical interpretation of errors in syllogistic reasoning. *J. Psychol.,* 1957, **43**, 341–344.

RUSSELL, B. The axiom of infinity. *Hibbert J.,* 1904, **2**, 809–812.

SCHILLER, F. C. S. *Logic for use.* New York: Harcourt, Brace, 1930.

SPINOZA, B. *Selections.* (Ed. by J. Wild) New York: Scribner, 1930.

VON DOMARUS, E. The specific laws of logic in schizophrenia. In J. S. Kasanin (Ed.), *Language and thought in schizophrenia.* Berkeley: Univer. California Press, 1944. Pp. 104–113.

WERTHEIMER, M. The syllogism and productive thinking. In W. D. Ellis (Ed.), *A source book of gestalt psychology.* London: Kegan Paul, Trench, Trubner, 1938. Selection 23.

WERTHEIMER, M. *Productive thinking.* (Enlarged ed.) New York: Harper, 1959.

(Received December 13, 1960)

ERRATA

In an article by Joan E. Jones, "All-or-None versus Incremental Learning," published in *Psychological Review,* 1962, **69**, 156–160, errors appeared in the numbering of experiments. On page 159 in Table 1, and in the last paragraph of the text on page 159 where Table 1 is discussed, where Experiment III appears it should be read as Experiment I.

In the table of contents on the front cover of the May 1962 issue, tiguity and Reinforcement in Relation to CS-UCS Intervals in Classical Aversive Conditioning."

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by BORIS SEMEONOFF

Vol. 53 Part 2 May, 1962 20s. net

ROBERT T. GREEN. The absolute threshold of electric shock.

J. L. ZAJAC. Some investigations of perception of movement and related depth phenomena.

RILEY W. GARDNER and ROBERT I. LONG. Control, defence and centration effect: a study of scanning behaviour.

RAY OVER. Stimulus wavelength variation and size and distance judgements.

S. RACHMAN. Disinhibition and the reminiscence effect in a motor learning task.

ARTHUR R. JENSEN. Is the serial-position curve invariant?

H. M. B. HURWITZ. Some properties of behaviour under fixed ratio and counting schedules.

K. LOVELL, B. MITCHELL and I. R. EVERETT. An experimental study of the growth of some logical structures.

D. E. BROADBENT and ALASTAIR HERON. Effects of a subsidiary task on performance involving immediate memory by younger and older men.

PUBLICATIONS RECENTLY RECEIVED.

The subscription price per volume, payable in advance, is 60s. net (post free).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS

Bentley House, Euston Road, London, N. W. 1

CONTEMPORARY PSYCHOLOGY

. . . APA's journal of book reviews

- Critical reviews of the most recent books in psychology and related areas
- The Editor's page: comments on news in the publishing world, on criticism, on literary style
- A letters-to-the-editor section
- Instructional Media: reviews and comments on films, TV programs, workbooks, teaching machines

AMERICAN PSYCHOLOGICAL ASSOCIATION

1333 Sixteenth St., N.W.
Washington 6, D. C.